

BS EN 13710:2011



BSI Standards Publication

European Ordering Rules — Ordering of characters from Latin, Greek, Cyrillic, Georgian and Armenian scripts

bsi.

...making excellence a habit.™

National foreword

This British Standard is the UK implementation of EN 13710:2011. It supersedes DD ENV 13710:2001 which is withdrawn.

The UK participation in its preparation was entrusted to Technical Committee ICT/-/1, Information systems co-ordination.

A list of organizations represented on this committee can be obtained on request to its secretary.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

© BSI 2011

ISBN 978 0 580 69139 3

ICS 01.140.20; 35.040

Compliance with a British Standard cannot confer immunity from legal obligations.

This British Standard was published under the authority of the Standards Policy and Strategy Committee on 31 March 2011.

Amendments issued since publication

Date	Text affected
------	---------------

English Version

European Ordering Rules - Ordering of characters from Latin,
Greek, Cyrillic, Georgian and Armenian scriptsRègles de classement européen - Classement des
caractères latins, grecs, cyrilliques, géorgiens et arméniensEuropäische Sortierregeln - Sortierung von lateinischen,
griechischen, kyrillischen, georgischen und armenischen
Schriftzeichen

This European Standard was approved by CEN on 5 February 2011.

CEN members are bound to comply with the CEN/CENELEC Internal Regulations which stipulate the conditions for giving this European Standard the status of a national standard without any alteration. Up-to-date lists and bibliographical references concerning such national standards may be obtained on application to the CEN-CENELEC Management Centre or to any CEN member.

This European Standard exists in three official versions (English, French, German). A version in any other language made by translation under the responsibility of a CEN member into its own language and notified to the CEN-CENELEC Management Centre has the same status as the official versions.

CEN members are the national standards bodies of Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland and United Kingdom.

EUROPEAN COMMITTEE FOR STANDARDIZATION
COMITÉ EUROPÉEN DE NORMALISATION
EUROPÄISCHES KOMITEE FÜR NORMUNG**Management Centre: Avenue Marnix 17, B-1000 Brussels**

Contents

Page

Foreword.....	3
Introduction	4
1 Scope	5
2 Normative references	5
3 Terms and definitions	5
4 Conformance	6
5 Tailorability	6
6 EOB Delta Table	7
Annex A (informative) Principles behind the European Ordering Rules	14
Annex B (informative) Word-by-word ordering	34
Annex C (informative) Ordering by position and by style	35
Annex D (informative) Mixed-script ordering with one predominant script	36
Annex E (informative) Defining National Deltas based on the EOB	37
Annex F (informative) Modern European Scripts / MES	42
Annex G (informative) EOB Delta in LDML Syntax	44
Bibliography	51

Foreword

This document (EN 13710:2011) has been prepared by Technical Committee CEN/TC 304 "Information and communications technologies - European localization requirements", the secretariat of which is held by DIN.

This European Standard shall be given the status of a national standard, either by publication of an identical text or by endorsement, at the latest by September 2011, and conflicting national standards shall be withdrawn at the latest by September 2011.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. CEN [and/or CENELEC] shall not be held responsible for identifying any or all such patent rights.

This document supersedes ENV 13710:2000 and CR 14400:2001.

This edition incorporates the following main changes made with respect to the previous edition of ENV 13710:2000 mainly as follows:

- a) ENV 13170:2000 and CR 14400:2001 have been consolidated;
- b) the document has been partly revised and has been brought up to date.

According to the CEN/CENELEC Internal Regulations, the national standards organizations of the following countries are bound to implement this European Standard: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland and United Kingdom.

Introduction

This European Standard provides rules for ordering multilingual lists into a well-defined and intuitive sequence. These rules are intended for data from different European languages that must be brought into a predictable order that makes it easy for users from multiple cultural backgrounds to find information. At the same time the standard is a basis for the definition of language-specific profiles taking the rules of a given language community into account at the same time as the total pan-European character set in a consistent, pan-European manner.

The rules have been tested and widely adopted in two predecessor specifications, ENV 13710:2000 *European Ordering Rules — Ordering of characters from the Latin, Greek and Cyrillic scripts* and its companion and extension specification CR 14400:2001 *European Ordering Rules - Ordering for Latin, Greek, Cyrillic, Georgian and Armenian scripts*. This European Standard consolidates these two documents into one technically largely upwards compatible standard.

This European Standard caters for two different target groups, software implementers on the one hand and users of ordering applications on the other.

Software implementers need unambiguous, machine-processable guidelines, which can readily be loaded into existing and future ordering applications. This goal can best be achieved by defining a european default ordering table in the syntaxes of two internationally relevant specifications in the field:

- the international ordering standard ISO/IEC 14651:2007 International string ordering and comparison — Method for comparing character strings and description of the common template tailorable ordering, of which the present standard is a “delta” (Clause 6);
- the Locale Data Markup Language (LDML) which is used to capture (amongst others) the collation data for the Unicode Common Locale Data Repository (CLDR). CLDR builds on the Unicode Collation Algorithm (UCA) which is technically a profile of ISO/IEC 14651:2007 (informative Annex G).

Users with no specific ICT background, however, need an explanation of the principles in a form more in line with existing national ordering standards or relevant practice. Tailoring tables can be difficult to read for human readers, so an explanation of the principles behind that table is given in the informative annexes. Users not familiar with the formal syntax of the tailoring table are advised to consult those annexes first.

The normative main part of this European Standard specifies letter-by-letter ordering of character strings. Informative Annex A presents equivalent information in a more human-oriented 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. Informative Annex E gives guidance on the use of this European Standard as the basis for expressing national deltas. Informative Annex F lists the underlying character repertoire for ease of reference. Informative Annex G expresses the formal delta in the LDML syntax.

Following the practice of ISO/IEC 14651 characters are referenced as UXXXX where X stands for any hexadecimal digit and refers to the code position of that character in ISO/IEC 10646. This convention is used throughout this European Standard.

1 Scope

This European Standard specifies the order between two character strings composed of characters from the *Modern European Scripts* (MES) collection of ISO/IEC 10646:2003 or subsets of it.

NOTE Collection 283 *Modern European Scripts* (MES) of ISO/IEC 10646:2003 was originally specified in CEN Workshop Agreement 13873:2000 *Multilingual European Subsets of ISO/IEC 10646* as *Multilingual European Subset Number 3* and was subsequently incorporated as a collection in Annex A of ISO/IEC 10646:2003 alongside its sister collections MES-1 and MES-2.

The ordering rules specified in this European Standard are only applicable for lists of data in more than one European language and when this data is intended for a multicultural audience. They complement existing national standards or practices in the field.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 10646:2003, *Information technology — Universal Multi-Octet Coded Character Set (UCS)*

ISO 12199:2000, *Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet*

ISO/IEC 14651:2007, *Information technology — International string ordering and comparison — Method for comparing character strings and description of the common template tailorable ordering*

3 Terms and definitions

For the purposes of this document, the following terms and definitions of ISO/IEC 10646:2003 and of ISO/IEC 14651:2007 apply.

3.1

character

member of a set of elements used for the organisation, control, or representation of data
[ISO/IEC 10646:2003]

3.2

character string

sequence of characters considered as a single object
[ISO/IEC 14651:2007]

3.3

collating symbol

symbol used to specify weights assigned to a collating element
[ISO/IEC 14651:2007]

3.4

collating element

sequence of one or more characters that are considered a single entity for ordering
[ISO/IEC 14651:2007]

3.5

collation table

mapping from collating elements to weighting elements
[ISO/IEC 14651:2007]

NOTE A special collation table is the Common Template Table (CTT) used in Annex A of ISO/IEC 14651 to express the default mapping from collating elements to weighting elements.

3.6

delta

list of the differences between a given collation table and another one
[ISO/IEC 14651]

NOTE The given collation table, together with a given delta, forms a new collation table. Unless otherwise specified in the European Standard, the term "delta" always refers to differences from the Common Template Table as defined in ISO/IEC 14651.

3.7

ordering

process by which, given two strings, it is determined whether the first one is less than, equal to, or greater than the second one
[ISO/IEC 14651:2007]

3.8

sorting

presentation of information in a structured way

NOTE Sorting may include the subdivision of information by subject matters, e.g. by having several registers in a book, by splitting a phone book into several sections, one for each town that falls into its purview or by having multiple indices in a library. Ordering is in most circumstances an integral part of this procedure.

4 Conformance

In order to be conformant to this European Standard an application shall meet the requirements prescribed in ISO/IEC 14651:2007, Clause 6 and its Common Template Table ISO14651_2006_TABLE1 after the application of the EOR delta table specified in Clause 6 of this European Standard. An equivalent description of the resulting tailored table shall equally conform to this European Standard.

5 Tailorability

The European Ordering Rules defined in this European Standard can be taken as a default template which can be tailored to the needs of any European country in the manner specified by ISO/IEC 14651 (cf. also informative Annex E).

This European Standard is not meant to influence national standards or traditions in the field of ordering, its scope being the ordering of multilingual data. Nonetheless, national standards are encouraged to express their national ordering rules on this European Standard by declaring a formalized set of deviation rules ("delta"), as explained in Informative Annex E. This way, the respective ordering rules are automatically machine-processable and can be incorporated into international repositories of locale data, allowing for more widespread support of national ordering standards across software products.

6 EOR Delta Table

NOTE For the syntax of the table please consult ISO/IEC 14651.

```
%% EOR's EORDeltaTable
%
%% European Ordering Rules.
%
% EOR delta for MES-3 from ISO/IEC 14651:2007's CTT (ISO14651_2006_TABLE1_en.txt).
%
% EOR uses four levels for comparison

reorder-after <BASE> % Introduce the LIG weight.
collating-symbol <LIG>
<LIG>
reorder-end

reorder-after <VRNT3> %Introduce more variants
collating-symbol <VRNT4>
collating-symbol <VRNT5>
collating-symbol <VRNT6>
collating-symbol <VRNT7>
collating-symbol <VRNT8>
collating-symbol <VRNT9>
<VRNT4>
<VRNT5>
<VRNT6>
<VRNT7>
<VRNT8>
<VRNT9>
reorder-end

reorder-after <S0584> %Introduce a weight for U0587 ARMENIAN SMALL LIGATURE ECH YIWN
collating-symbol <ECH-YIWN>
<ECH-YIWN>
reorder-end

reorder-after <SFFFF>

order_start forward;forward;forward;forward

% Non-alphanumeric characters (including some modifier letters):

% The DRACHMA SIGN is already in ISO14651_2006 ignorable on levels 1-3
<U0024> IGNORE:IGNORE:IGNORE;<U0024> % DOLLAR SIGN
<U00A2> IGNORE:IGNORE:IGNORE;<U00A2> % CENT SIGN
<U00A3> IGNORE:IGNORE:IGNORE;<U00A3> % POUND SIGN
<U00A4> IGNORE:IGNORE:IGNORE;<U00A4> % CURRENCY SIGN
<U00A5> IGNORE:IGNORE:IGNORE;<U00A5> % YEN SIGN
<U20A0> IGNORE:IGNORE:IGNORE;<U20A0> % EURO-CURRENCY SIGN
```

<U20A1> IGNORE:IGNORE:IGNORE:<U20A1> % COLON SIGN
<U20A2> IGNORE:IGNORE:IGNORE:<U20A2> % CRUZEIRO SIGN
<U20A3> IGNORE:IGNORE:IGNORE:<U20A3> % FRENCH FRANC SIGN
<U20A4> IGNORE:IGNORE:IGNORE:<U20A4> % LIRA SIGN
<U20A5> IGNORE:IGNORE:IGNORE:<U20A5> % MILL SIGN
<U20A6> IGNORE:IGNORE:IGNORE:<U20A6> % NAIRA SIGN
<U20A7> IGNORE:IGNORE:IGNORE:<U20A7> % PESETA SIGN
<U20A8> IGNORE:IGNORE:IGNORE:<U20A8> % RUPEE SIGN
<U20A9> IGNORE:IGNORE:IGNORE:<U20A9> % WON SIGN
<U20AA> IGNORE:IGNORE:IGNORE:<U20AA> % NEW SHEQEL SIGN
<U20AB> IGNORE:IGNORE:IGNORE:<U20AB> % DONG SIGN
<U20AC> IGNORE:IGNORE:IGNORE:<U20AC> % EURO SIGN
<U20AD> IGNORE:IGNORE:IGNORE:<U20AD> % KIP SIGN
<U20AE> IGNORE:IGNORE:IGNORE:<U20AE> % TUGRIK SIGN
<U20AF> IGNORE:IGNORE:IGNORE:<U20AF> % DRACHMA SIGN
<U20B0> IGNORE:IGNORE:IGNORE:<U20B0> % GERMAN PENNY SIGN
<U20B1> IGNORE:IGNORE:IGNORE:<U20B1> % PESO SIGN
<U20B2> IGNORE:IGNORE:IGNORE:<U20B2> % GUARANI SIGN
<U20B3> IGNORE:IGNORE:IGNORE:<U20B3> % AUSTRAL SIGN
<U20B4> IGNORE:IGNORE:IGNORE:<U20B4> % HRYVNIA SIGN
<U20B5> IGNORE:IGNORE:IGNORE:<U20B5> % CEDI SIGN

% Modifier letters that are not ignorable in ISO14651_2006_TABLE1_en.txt

<U02B0> IGNORE:IGNORE:IGNORE:<U02B0> % MODIFIER LETTER SMALL H
<U02B1> IGNORE:IGNORE:IGNORE:<U02B1> % MODIFIER LETTER SMALL H WITH HOOK
<U02B2> IGNORE:IGNORE:IGNORE:<U02B2> % MODIFIER LETTER SMALL J
<U02B3> IGNORE:IGNORE:IGNORE:<U02B3> % MODIFIER LETTER SMALL R
<U02B4> IGNORE:IGNORE:IGNORE:<U02B4> % MODIFIER LETTER SMALL TURNED R
<U02B5> IGNORE:IGNORE:IGNORE:<U02B5> % MODIFIER LETTER SMALL TURNED R WITH HOOK
<U02B6> IGNORE:IGNORE:IGNORE:<U02B6> % MODIFIER LETTER SMALL CAPITAL INVERTED R
<U02B7> IGNORE:IGNORE:IGNORE:<U02B7> % MODIFIER LETTER SMALL W
<U02B8> IGNORE:IGNORE:IGNORE:<U02B8> % MODIFIER LETTER SMALL Y
<U02BB> IGNORE:IGNORE:IGNORE:<U02BB> % MODIFIER LETTER TURNED COMMA
<U02BC> IGNORE:IGNORE:IGNORE:<U02BC> % MODIFIER LETTER APOSTROPHE
<U02BD> IGNORE:IGNORE:IGNORE:<U02BD> % MODIFIER LETTER REVERSED COMMA
<U02BE> IGNORE:IGNORE:IGNORE:<U02BE> % MODIFIER LETTER RIGHT HALF RING
<U02BF> IGNORE:IGNORE:IGNORE:<U02BF> % MODIFIER LETTER LEFT HALF RING
<U02C0> IGNORE:IGNORE:IGNORE:<U02C0> % MODIFIER LETTER GLOTTAL STOP
<U02C1> IGNORE:IGNORE:IGNORE:<U02C1> % MODIFIER LETTER REVERSED GLOTTAL STOP
<U02D0> IGNORE:IGNORE:IGNORE:<U02D0> % MODIFIER LETTER TRIANGULAR COLON
<U02D1> IGNORE:IGNORE:IGNORE:<U02D1> % MODIFIER LETTER HALF TRIANGULAR COLON
<U02E0> IGNORE:IGNORE:IGNORE:<U02E0> % MODIFIER LETTER SMALL GAMMA
<U02E1> IGNORE:IGNORE:IGNORE:<U02E1> % MODIFIER LETTER SMALL L
<U02E2> IGNORE:IGNORE:IGNORE:<U02E2> % MODIFIER LETTER SMALL S
<U02E4> IGNORE:IGNORE:IGNORE:<U02E4> % MODIFIER LETTER SMALL REVERSED GLOTTAL STOP
<U02EE> IGNORE:IGNORE:IGNORE:<U02EE> % MODIFIER LETTER DOUBLE APOSTROPHE

<U0294> IGNORE:IGNORE:IGNORE:<U0294> % LATIN LETTER GLOTTAL STOP
<U0295> IGNORE:IGNORE:IGNORE:<U0295> % LATIN LETTER PHARYNGEAL VOICED FRICATIVE
<U0296> IGNORE:IGNORE:IGNORE:<U0296> % LATIN LETTER INVERTED GLOTTAL STOP
<U0298> IGNORE:IGNORE:IGNORE:<U0298> % LATIN LETTER BILABIAL CLICK
<U02A1> IGNORE:IGNORE:IGNORE:<U02A1> % LATIN LETTER GLOTTAL STOP WITH STROKE

<U02A2> IGNORE:IGNORE:IGNORE:<U02A2> % LATIN LETTER REVERSED GLOTTAL STOP WITH STROKE

%%

% Latin

% Almost all changes here result from CEN/TC304's resolution

% for the Latin script part of the Modern European Scripts / MES-3 to

% treat only the letters a to z and thorn as distinct on the first

% level and treat other combinations as variants or ligatures

<U0250> <S0061>;<BASE><VRNT1>;<MIN><MIN>;<U0250> % LATIN SMALL LETTER TURNED A

<U0251> <S0061>;<BASE><VRNT2>;<MIN><MIN>;<U0251> % LATIN SMALL LETTER ALPHA

<U0252> <S0061>;<BASE><VRNT3>;<MIN><MIN>;<U0252> % LATIN SMALL LETTER TURNED ALPHA

<U0299> <S0062>;<BASE><VRNT1>;<MIN><MIN>;<U0299> % LATIN LETTER SMALL CAPITAL B

<U0180> <S0062>;<BASE><VRNT2>;<MIN><MIN>;<U0180> % LATIN SMALL LETTER B WITH STROKE

<U0243> <S0062>;<BASE><VRNT2>;<CAP><MIN>;<U0243> % LATIN CAPITAL LETTER B WITH STROKE

<U0253> <S0062>;<BASE><VRNT3>;<MIN><MIN>;<U0253> % LATIN SMALL LETTER B WITH HOOK

<U0181> <S0062>;<BASE><VRNT3>;<CAP><MIN>;<U0181> % LATIN CAPITAL LETTER B WITH HOOK

<U0183> <S0062>;<BASE><VRNT4>;<MIN><MIN>;<U0183> % LATIN SMALL LETTER B WITH TOPBAR

<U0182> <S0062>;<BASE><VRNT4>;<CAP><MIN>;<U0182> % LATIN CAPITAL LETTER B WITH TOPBAR

<U0188> <S0063>;<BASE><VRNT1>;<MIN><MIN>;<U0188> % LATIN SMALL LETTER C WITH HOOK

<U0187> <S0063>;<BASE><VRNT1>;<CAP><MIN>;<U0187> % LATIN CAPITAL LETTER C WITH HOOK

<U0255> <S0063>;<BASE><VRNT2>;<MIN><MIN>;<U0255> % LATIN SMALL LETTER C WITH CURL

<U0297> <S0063>;<BASE><VRNT3>;<MIN><MIN>;<U0297> % LATIN LETTER STRETCHED C

% <VRNT1> is used for U00F0 LATIN SMALL LETTER ETH (already in CTT)

<U0256> <S0064>;<BASE><VRNT2>;<MIN><MIN>;<U0256> % LATIN SMALL LETTER D WITH TAIL

<U0189> <S0064>;<BASE><VRNT2>;<CAP><MIN>;<U0189> % LATIN CAPITAL LETTER AFRICAN D

<U0257> <S0064>;<BASE><VRNT3>;<MIN><MIN>;<U0257> % LATIN SMALL LETTER D WITH HOOK

<U018A> <S0064>;<BASE><VRNT3>;<CAP><MIN>;<U018A> % LATIN CAPITAL LETTER D WITH HOOK

<U018C> <S0064>;<BASE><VRNT4>;<MIN><MIN>;<U018C> % LATIN SMALL LETTER D WITH TOPBAR

<U018B> <S0064>;<BASE><VRNT4>;<CAP><MIN>;<U018B> % LATIN CAPITAL LETTER D WITH TOPBAR

<U0221> <S0064>;<BASE><VRNT5>;<MIN><MIN>;<U0221> % LATIN SMALL LETTER D WITH CURL

<U018D> <S0064>;<BASE><VRNT6>;<MIN><MIN>;<U018D> % LATIN SMALL LETTER TURNED DELTA

<U02A5> "<S0064><S007A>;<BASE><BASE><VRNT4>;<COMPAT><COMPAT><COMPAT>;<U02A5> % LATIN SMALL LETTER DZ DIGRAPH WITH CURL

<U02A4> "<S0064><S007A>;<BASE><BASE><VRNT5>;<COMPAT><COMPAT><COMPAT>;<U02A4> % LATIN SMALL LETTER DEZH DIGRAPH

<U01DD> <S0065>;<BASE><VRNT1>;<MIN><MIN>;<U01DD> % LATIN SMALL LETTER TURNED E

<U018E> <S0065>;<BASE><VRNT1>;<CAP><MIN>;<U018E> % LATIN CAPITAL LETTER REVERSED E

<U0259> <S0065>;<BASE><VRNT2>;<MIN><MIN>;<U0259> % LATIN SMALL LETTER SCHWA

<U018F> <S0065>;<BASE><VRNT2>;<CAP><MIN>;<U018F> % LATIN CAPITAL LETTER SCHWA

<U025B> <S0065>;<BASE><VRNT3>;<MIN><MIN>;<U025B> % LATIN SMALL LETTER OPEN E

<U0190> <S0065>;<BASE><VRNT3>;<CAP><MIN>;<U0190> % LATIN CAPITAL LETTER OPEN E

<U0258> <S0065>;<BASE><VRNT4>;<MIN><MIN>;<U0258> % LATIN SMALL LETTER REVERSED E

<U025A> <S0065>;<BASE><VRNT5>;<MIN><MIN>;<U025A> % LATIN SMALL LETTER SCHWA WITH HOOK

<U025C> <S0065>;<BASE><VRNT6>;<MIN><MIN>;<U025C> % LATIN SMALL LETTER REVERSED OPEN E

<U025D> <S0065>;<BASE><VRNT7>;<MIN><MIN>;<U025D> % LATIN SMALL LETTER REVERSED OPEN E WITH HOOK

<U025E> <S0065>;<BASE><VRNT8>;<MIN><MIN>;<U025E> % LATIN SMALL LETTER CLOSED REVERSED OPEN E

<U029A> <S0065>;" <BASE><VRNT9>;" <MIN><MIN>;" <U029A> % LATIN SMALL LETTER CLOSED OPEN E

<U0192> <S0066>;" <BASE><VRNT1>;" <MIN><MIN>;" <U0192> % LATIN SMALL LETTER F WITH HOOK
<U0191> <S0066>;" <BASE><VRNT1>;" <CAP><MIN>;" <U0191> % LATIN CAPITAL LETTER F WITH HOOK

<U0261> <S0067>;" <BASE><VRNT1>;" <MIN><MIN>;" <U0261> % LATIN SMALL LETTER SCRIPT G
<U0262> <S0067>;" <BASE><VRNT2>;" <MIN><MIN>;" <U0262> % LATIN LETTER SMALL CAPITAL G
<U01E5> <S0067>;" <BASE><VRNT3>;" <MIN><MIN>;" <U01E5> % LATIN SMALL LETTER G WITH STROKE
<U01E4> <S0067>;" <BASE><VRNT3>;" <CAP><MIN>;" <U01E4> % LATIN CAPITAL LETTER G WITH STROKE
<U0260> <S0067>;" <BASE><VRNT4>;" <MIN><MIN>;" <U0260> % LATIN SMALL LETTER G WITH HOOK
<U0193> <S0067>;" <BASE><VRNT4>;" <CAP><MIN>;" <U0193> % LATIN CAPITAL LETTER G WITH HOOK
<U029B> <S0067>;" <BASE><VRNT5>;" <MIN><MIN>;" <U029B> % LATIN LETTER SMALL CAPITAL G WITH HOOK
<U0263> <S0067>;" <BASE><VRNT6>;" <MIN><MIN>;" <U0263> % LATIN SMALL LETTER GAMMA
<U0194> <S0067>;" <BASE><VRNT6>;" <CAP><MIN>;" <U0194> % LATIN CAPITAL LETTER GAMMA
<U0264> <S0067>;" <BASE><VRNT7>;" <MIN><MIN>;" <U0264> % LATIN SMALL LETTER RAMS HORN
<U01A3> <S0067>;" <BASE><VRNT8>;" <MIN><MIN>;" <U01A3> % LATIN SMALL LETTER OI
<U01A2> <S0067>;" <BASE><VRNT8>;" <CAP><MIN>;" <U01A2> % LATIN CAPITAL LETTER OI

<U029C> <S0068>;" <BASE><VRNT1>;" <MIN><MIN>;" <U029C> % LATIN LETTER SMALL CAPITAL H
<U0266> <S0068>;" <BASE><VRNT2>;" <MIN><MIN>;" <U0266> % LATIN SMALL LETTER H WITH HOOK
<U0267> <S0068>;" <BASE><VRNT3>;" <MIN><MIN>;" <U0267> % LATIN SMALL LETTER HENG WITH HOOK
<U0265> <S0068>;" <BASE><VRNT4>;" <MIN><MIN>;" <U0265> % LATIN SMALL LETTER TURNED H
<U02AE> <S0068>;" <BASE><VRNT5>;" <MIN><MIN>;" <U02AE> % LATIN SMALL LETTER TURNED H WITH FISHHOOK
<U02AF> <S0068>;" <BASE><VRNT6>;" <MIN><MIN>;" <U02AF> % LATIN SMALL LETTER TURNED H WITH FISHHOOK AND TAIL

<U0195> " <S0068><S0076>;" <BASE><BASE>;" <MIN><MIN>;" <U0195> % LATIN SMALL LETTER HV
<U01F6> " <S0068><S0076>;" <BASE><BASE>;" <CAP><MIN>;" <U01F6> % LATIN CAPITAL LETTER HWAIR

<U0131> <S0069>;" <BASE><VRNT1>;" <MIN><MIN>;" <U0131> % LATIN SMALL LETTER DOTLESS I
<U026A> <S0069>;" <BASE><VRNT2>;" <MIN><MIN>;" <U026A> % LATIN LETTER SMALL CAPITAL I
<U0268> <S0069>;" <BASE><VRNT3>;" <MIN><MIN>;" <U0268> % LATIN SMALL LETTER I WITH STROKE
<U0197> <S0069>;" <BASE><VRNT3>;" <CAP><MIN>;" <U0197> % LATIN CAPITAL LETTER I WITH STROKE
<U0269> <S0069>;" <BASE><VRNT4>;" <MIN><MIN>;" <U0269> % LATIN SMALL LETTER IOTA
<U0196> <S0069>;" <BASE><VRNT4>;" <CAP><MIN>;" <U0196> % LATIN CAPITAL LETTER IOTA

<U029D> <S006A>;" <BASE><VRNT1>;" <MIN><MIN>;" <U029D> % LATIN SMALL LETTER J WITH CROSSED-TAIL
<U025F> <S006A>;" <BASE><VRNT2>;" <MIN><MIN>;" <U025F> % LATIN SMALL LETTER DOTLESS J WITH STROKE
<U0284> <S006A>;" <BASE><VRNT3>;" <MIN><MIN>;" <U0284> % LATIN SMALL LETTER DOTLESS J WITH STROKE AND HOOK

<U0199> <S006B>;" <BASE><VRNT1>;" <MIN><MIN>;" <U0199> % LATIN SMALL LETTER K WITH HOOK
<U0198> <S006B>;" <BASE><VRNT1>;" <CAP><MIN>;" <U0198> % LATIN CAPITAL LETTER K WITH HOOK
<U0138> <S006B>;" <BASE><VRNT2>;" <MIN><MIN>;" <U0138> % LATIN SMALL LETTER KRA
<U029E> <S006B>;" <BASE><VRNT3>;" <MIN><MIN>;" <U029E> % LATIN SMALL LETTER TURNED K

% <VRNT1> is used for U0140 LATIN SMALL LETTER L WITH MIDDLE DOT (already in CTT)

<U029F> <S006C>;" <BASE><VRNT2>;" <MIN><MIN>;" <U029F> % LATIN LETTER SMALL CAPITAL L
<U019A> <S006C>;" <BASE><VRNT3>;" <MIN><MIN>;" <U019A> % LATIN SMALL LETTER L WITH BAR
<U023D> <S006C>;" <BASE><VRNT3>;" <CAP><MIN>;" <U023D> % LATIN CAPITAL LETTER L WITH BAR
<U026B> <S006C>;" <BASE><VRNT4>;" <MIN><MIN>;" <U026B> % LATIN SMALL LETTER L WITH MIDDLE TILDE
<U026C> <S006C>;" <BASE><VRNT5>;" <MIN><MIN>;" <U026C> % LATIN SMALL LETTER L WITH BELT
<U026D> <S006C>;" <BASE><VRNT6>;" <MIN><MIN>;" <U026D> % LATIN SMALL LETTER L WITH RETROFLEX HOOK
<U0234> <S006C>;" <BASE><VRNT7>;" <MIN><MIN>;" <U0234> % LATIN SMALL LETTER L WITH CURL

<U019B> <S006C>;" <BASE><VRNT8>;" <MIN><MIN>; <U019B> % LATIN SMALL LETTER LAMBDA WITH STROKE

<U026E> " <S006C><S007A>;" <BASE><BASE><VRNT5>;" <MIN><MIN><MIN>; <U026E> % LATIN SMALL LETTER LEZH

<U0271> <S006D>;" <BASE><VRNT1>;" <MIN><MIN>; <U0271> % LATIN SMALL LETTER M WITH HOOK

<U026F> <S006D>;" <BASE><VRNT2>;" <MIN><MIN>; <U026F> % LATIN SMALL LETTER TURNED M

<U019C> <S006D>;" <BASE><VRNT2>;" <CAP><MIN>; <U019C> % LATIN CAPITAL LETTER TURNED M

<U0270> <S006D>;" <BASE><VRNT3>;" <MIN><MIN>; <U0270> % LATIN SMALL LETTER TURNED M WITH LONG LEG

<U0149> <S006E>;" <BASE><VRNT1>;" <MIN><MIN>; <U0149> % LATIN SMALL LETTER N PRECEDED BY APOSTROPHE

<U0274> <S006E>;" <BASE><VRNT2>;" <MIN><MIN>; <U0274> % LATIN LETTER SMALL CAPITAL N

<U0272> <S006E>;" <BASE><VRNT3>;" <MIN><MIN>; <U0272> % LATIN SMALL LETTER N WITH LEFT HOOK

<U019D> <S006E>;" <BASE><VRNT3>;" <CAP><MIN>; <U019D> % LATIN CAPITAL LETTER N WITH LEFT HOOK

<U019E> <S006E>;" <BASE><VRNT4>;" <MIN><MIN>; <U019E> % LATIN SMALL LETTER N WITH LONG RIGHT LEG

<U0220> <S006E>;" <BASE><VRNT4>;" <CAP><MIN>; <U0220> % LATIN CAPITAL LETTER N WITH LONG RIGHT LEG

<U0273> <S006E>;" <BASE><VRNT5>;" <MIN><MIN>; <U0273> % LATIN SMALL LETTER N WITH RETROFLEX HOOK

<U0235> <S006E>;" <BASE><VRNT6>;" <MIN><MIN>; <U0235> % LATIN SMALL LETTER N WITH CURL

<U014B> <S006E>;" <BASE><VRNT7>;" <MIN><MIN>; <U014B> % LATIN SMALL LETTER ENG

<U014A> <S006E>;" <BASE><VRNT7>;" <CAP><MIN>; <U014A> % LATIN CAPITAL LETTER ENG

% <VRNT1> is used for U0153 LATIN SMALL LIGATURE OE (already in CTT)

<U0254> <S006F>;" <BASE><VRNT2>;" <MIN><MIN>; <U0254> % LATIN SMALL LETTER OPEN O

<U0186> <S006F>;" <BASE><VRNT2>;" <CAP><MIN>; <U0186> % LATIN CAPITAL LETTER OPEN O

<U0275> <S006F>;" <BASE><VRNT3>;" <MIN><MIN>; <U0275> % LATIN SMALL LETTER BARRED O

<U019F> <S006F>;" <BASE><VRNT3>;" <CAP><MIN>; <U019F> % LATIN CAPITAL LETTER O WITH MIDDLE TILDE

<U0277> <S006F>;" <BASE><VRNT4>;" <MIN><MIN>; <U0277> % LATIN SMALL LETTER CLOSED OMEGA

<U0223> <S006F>;" <BASE><VRNT5>;" <MIN><MIN>; <U0223> % LATIN SMALL LETTER OU

<U0222> <S006F>;" <BASE><VRNT5>;" <CAP><MIN>; <U0222> % LATIN CAPITAL LETTER OU

<U0276> " <S006F><S0065>;" <BASE><VRNT1><BASE>;" <COMPAT><COMPAT><COMPAT>; <U0276> % LATIN LETTER SMALL CAPITAL OE

<U01A5> <S0070>;" <BASE><VRNT1>;" <MIN><MIN>; <U01A5> % LATIN SMALL LETTER P WITH HOOK

<U01A4> <S0070>;" <BASE><VRNT1>;" <CAP><MIN>; <U01A4> % LATIN CAPITAL LETTER P WITH HOOK

<U0278> <S0070>;" <BASE><VRNT2>;" <MIN><MIN>; <U0278> % LATIN SMALL LETTER PHI

<U02A0> <S0071>;" <BASE><VRNT1>;" <MIN><MIN>; <U02A0> % LATIN SMALL LETTER Q WITH HOOK

<U0280> <S0072>;" <BASE><VRNT1>;" <MIN><MIN>; <U01A6> % LATIN LETTER SMALL CAPITAL R

<U01A6> <S0072>;" <BASE><VRNT1>;" <CAP><MIN>; <U01A6> % LATIN LETTER YR

<U0279> <S0072>;" <BASE><VRNT2>;" <MIN><MIN>; <U0279> % LATIN SMALL LETTER TURNED R

<U027A> <S0072>;" <BASE><VRNT3>;" <MIN><MIN>; <U027A> % LATIN SMALL LETTER TURNED R WITH LONG LEG

<U027B> <S0072>;" <BASE><VRNT4>;" <MIN><MIN>; <U027B> % LATIN SMALL LETTER TURNED R WITH HOOK

<U027C> <S0072>;" <BASE><VRNT5>;" <MIN><MIN>; <U027C> % LATIN SMALL LETTER R WITH LONG LEG

<U027D> <S0072>;" <BASE><VRNT6>;" <MIN><MIN>; <U027D> % LATIN SMALL LETTER R WITH TAIL

<U027E> <S0072>;" <BASE><VRNT7>;" <MIN><MIN>; <U027E> % LATIN SMALL LETTER R WITH FISHHOOK

<U027F> <S0072>;" <BASE><VRNT8>;" <MIN><MIN>; <U027F> % LATIN SMALL LETTER REVERSED R WITH FISHHOOK

<U0281> <S0072>;" <BASE><VRNT9>;" <MIN><MIN>; <U0281> % LATIN LETTER SMALL CAPITAL INVERTED R

% <VRNT1> is used for U00DF LATIN SMALL LETTER SHARP S (already in CTT)

% <VRNT2> is used for U017F LATIN SMALL LETTER LONG S (already in CTT)

<U0282> <S0073>;" <BASE><VRNT3>;" <MIN><MIN>; <U0282> % LATIN SMALL LETTER S WITH HOOK

<U0283> <S0073>;" <BASE><VRNT4>;" <MIN><MIN>;" <U0283> % LATIN SMALL LETTER ESH
<U01A9> <S0073>;" <BASE><VRNT4>;" <CAP><MIN>;" <U01A9> % LATIN CAPITAL LETTER ESH
<U01AA> <S0073>;" <BASE><VRNT5>;" <MIN><MIN>;" <U01AA> % LATIN LETTER REVERSED ESH LOOP
<U0285> <S0073>;" <BASE><VRNT6>;" <MIN><MIN>;" <U0285> % LATIN SMALL LETTER SQUAT REVERSED ESH
<U0286> <S0073>;" <BASE><VRNT7>;" <MIN><MIN>;" <U0286> % LATIN SMALL LETTER ESH WITH CURL

<U0167> <S0074>;" <BASE><VRNT1>;" <MIN><MIN>;" <U0167> % LATIN SMALL LETTER T WITH STROKE
<U0166> <S0074>;" <BASE><VRNT1>;" <CAP><MIN>;" <U0166> % LATIN CAPITAL LETTER T WITH STROKE
<U01AB> <S0074>;" <BASE><VRNT2>;" <MIN><MIN>;" <U01AB> % LATIN SMALL LETTER T WITH PALATAL HOOK
<U01AD> <S0074>;" <BASE><VRNT3>;" <MIN><MIN>;" <U01AD> % LATIN SMALL LETTER T WITH HOOK
<U01AC> <S0074>;" <BASE><VRNT3>;" <CAP><MIN>;" <U01AC> % LATIN CAPITAL LETTER T WITH HOOK
<U0288> <S0074>;" <BASE><VRNT4>;" <MIN><MIN>;" <U0288> % LATIN SMALL LETTER T WITH RETROFLEX HOOK
<U01AE> <S0074>;" <BASE><VRNT4>;" <CAP><MIN>;" <U01AE> % LATIN CAPITAL LETTER T WITH RETROFLEX HOOK
<U0236> <S0074>;" <BASE><VRNT5>;" <MIN><MIN>;" <U0236> % LATIN SMALL LETTER T WITH CURL
<U0287> <S0074>;" <BASE><VRNT6>;" <MIN><MIN>;" <U0287> % LATIN SMALL LETTER TURNED T

<U02A8> " <S0074><S0063>;" <BASE><BASE><VRNT2>;" <COMPAT><COMPAT><COMPAT>;" <U02A8> % LATIN SMALL LETTER TC DIGRAPH WITH CURL

<U0289> <S0075>;" <BASE><VRNT1>;" <MIN><MIN>;" <U0289> % LATIN SMALL LETTER U BAR
<U0244> <S0075>;" <BASE><VRNT1>;" <CAP><MIN>;" <U0244> % LATIN CAPITAL LETTER U BAR
<U028A> <S0075>;" <BASE><VRNT2>;" <MIN><MIN>;" <U028A> % LATIN SMALL LETTER UPSILON
<U01B1> <S0075>;" <BASE><VRNT2>;" <CAP><MIN>;" <U01B1> % LATIN CAPITAL LETTER UPSILON

<U028B> <S0076>;" <BASE><VRNT1>;" <MIN><MIN>;" <U028B> % LATIN SMALL LETTER V WITH HOOK
<U01B2> <S0076>;" <BASE><VRNT1>;" <CAP><MIN>;" <U01B2> % LATIN CAPITAL LETTER V WITH HOOK
<U028C> <S0076>;" <BASE><VRNT2>;" <MIN><MIN>;" <U028C> % LATIN SMALL LETTER TURNED V
<U0245> <S0076>;" <BASE><VRNT2>;" <CAP><MIN>;" <U0245> % LATIN CAPITAL LETTER TURNED V

<U028D> <S0077>;" <BASE><VRNT1>;" <MIN><MIN>;" <U028D> % LATIN SMALL LETTER TURNED W
<U01BF> <S0077>;" <BASE><VRNT2>;" <MIN><MIN>;" <U01BF> % LATIN LETTER WYNN
<U01F7> <S0077>;" <BASE><VRNT2>;" <CAP><MIN>;" <U01F7> % LATIN CAPITAL LETTER WYNN

<U028F> <S0079>;" <BASE><VRNT1>;" <MIN><MIN>;" <U028F> % LATIN LETTER SMALL CAPITAL Y
<U01B4> <S0079>;" <BASE><VRNT2>;" <MIN><MIN>;" <U01B4> % LATIN SMALL LETTER Y WITH HOOK
<U01B3> <S0079>;" <BASE><VRNT2>;" <CAP><MIN>;" <U01B3> % LATIN CAPITAL LETTER Y WITH HOOK
<U028E> <S0079>;" <BASE><VRNT3>;" <MIN><MIN>;" <U028E> % LATIN SMALL LETTER TURNED Y
<U021D> <S0079>;" <BASE><VRNT4>;" <MIN><MIN>;" <U021D> % LATIN SMALL LETTER YOGH
<U021C> <S0079>;" <BASE><VRNT9>;" <CAP><MIN>;" <U021C> % LATIN CAPITAL LETTER YOGH

<U01B6> <S007A>;" <BASE><VRNT1>;" <MIN><MIN>;" <U01B6> % LATIN SMALL LETTER Z WITH STROKE
<U01B5> <S007A>;" <BASE><VRNT1>;" <CAP><MIN>;" <U01B5> % LATIN CAPITAL LETTER Z WITH STROKE
<U0225> <S007A>;" <BASE><VRNT2>;" <MIN><MIN>;" <U0225> % LATIN SMALL LETTER Z WITH HOOK
<U0224> <S007A>;" <BASE><VRNT2>;" <CAP><MIN>;" <U0224> % LATIN CAPITAL LETTER Z WITH HOOK
<U0290> <S007A>;" <BASE><VRNT3>;" <MIN><MIN>;" <U0290> % LATIN SMALL LETTER Z WITH RETROFLEX HOOK
<U0291> <S007A>;" <BASE><VRNT4>;" <MIN><MIN>;" <U0291> % LATIN SMALL LETTER Z WITH CURL
<U0292> <S007A>;" <BASE><VRNT5>;" <MIN><MIN>;" <U0292> % LATIN SMALL LETTER EZH
<U01B7> <S007A>;" <BASE><VRNT5>;" <CAP><MIN>;" <U01B7> % LATIN CAPITAL LETTER EZH
<U01EF> <S007A>;" <BASE><VRNT5><CARON>;" <MIN><MIN><MIN>;" <U01EF> % LATIN SMALL LETTER EZH WITH CARON
<U01EE> <S007A>;" <BASE><VRNT5><CARON>;" <CAP><MIN><MIN>;" <U01EE> % LATIN CAPITAL LETTER EZH WITH CARON
<U01B9> <S007A>;" <BASE><VRNT6>;" <MIN><MIN>;" <U01B9> % LATIN SMALL LETTER EZH REVERSED

<U01B8> <S007A>:"<BASE><VRNT6>";"<CAP><MIN>";<U01B8> % LATIN CAPITAL LETTER EZH REVERSED
<U01BA> <S007A>:"<BASE><VRNT7>";"<MIN><MIN>";<U01BA> % LATIN SMALL LETTER EZH WITH TAIL
<U0293> <S007A>:"<BASE><VRNT8>";"<MIN><MIN>";<U0293> % LATIN SMALL LETTER EZH WITH CURL

% Greek

% ISO14651_2006_TABLE1_en.txt now contains the tailorings of CR 14400 in its CTT

% Full conformance with GOST requirements for Cyrillic letters

<U0453> <S0452>:"<BASE><VRNT1>";"<MIN><MIN>";<U0453> % CYRILLIC SMALL LETTER GJE
<U0403> <S0452>:"<BASE><VRNT1>";"<CAP><MIN>";<U0403> % CYRILLIC CAPITAL LETTER GJE

<U045C> <S045B>:"<BASE><VRNT1>";"<MIN><MIN>";<U045C> % CYRILLIC SMALL LETTER KJE
<U040C> <S045B>:"<BASE><VRNT1>";"<CAP><MIN>";<U040C> % CYRILLIC CAPITAL LETTER KJE

% Georgian: Identical to ISO14651_2006_TABLE1_en.txt

% Armenian:

<U0587> <ECH-YIWN>;<BASE>;<CAP>;<U0587> % ARMENIAN SMALL LIGATURE ECH YIWN

reorder-end %% for EOR's EORDeltaTable

Annex A (informative)

Principles behind the European Ordering Rules

A.0 Introduction

This annex aims to present the information inherent in Clause 6 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 interpretation of the body of this European Standard.

A.1 Terms and definitions

For the purpose of this annex, the following terms and definitions apply in addition to those in the body of this European Standard (see Clause 3).

A.1.1

digit

any of the characters

0 (U0030), 1 (U0031), 2 (U0032), 3 (U0033), 4 (U0034), 5 (U0035), 6 (U0036), 7 (U0037),
8 (U0038), 9 (U0039)

A.1.2

letter

character used to represent (either alone or in combination) sounds or sequences of sounds of a natural language in writing

NOTE Here equivalent to all characters of the *Multilingual European Subset No 3* whose name contains one of the words LETTER or LIGATURE.

A.1.3

first level letter

character that is a member of the following list of letters:

Latin script:

a (U0061), A (U0041), b (U0062), B (U0042), c (U0063), C (U0043), d (U0064), D (U0044),
e (U0065), E (U0045), f (U0066), F (U0046), g (U0067), G (U0047), h (U0068), H (U0048),
i (U0069), I (U0049), j (U006A), J (U004A), k (U006B), K (U004B), l (U006C), L (U004C),
m (U006D), M (U004D), n (U006E), N (U004E), o (U006F), O (U004F), p (U0070), P (U0050),
q (U0071), Q (U0051), r (U0072), R (U0052), s (U0073), S (U0053), t (U0074), T (U0054),
u (U0075), U (U0055), v (U0076), V (U0056), w (U0077), W (U0057), x (U0078), X (U0058),

γ (U0079), Υ (U0059), ζ (U007A), Ζ (U005A), þ (U00FE), Þ (U00DE)

Greek script:

α (U03B1), Α (U0391), β (U03B2), Β (U0392), γ (U03B3), Γ (U0393), δ (U03B4), Δ (U0394),
ε (U03B5), Ε (U0395), Ϝ (U03DC), Ϛ (U03DA), Ϝ (U03B6), Ζ (U0396), η (U03B7), Η (U0397),
θ (U03B8), Θ (U0398), ι (U03B9), Ι (U0399), κ (U03BA), Κ (U039A), λ (U03BB), Λ (U039B),
μ (U03BC), Μ (U039C), ν (U03BD), Ν (U039D), ξ (U03BE), Ξ (U039E), ο (U03BF), Ο (U039F),
π (U03C0), Π (U03A0), Ϸ (U03DE), ϱ (U03C1), Ρ (U03A1), σ (U03C3), Σ (U03A3), τ (U03C4),
Τ (U03A4), υ (U03C5), Υ (U03A5), φ (U03C6), Φ (U03A6), χ (U03C7), Χ (U03A7), ψ (U03C8),
Ψ (U03A8), ω (U03C9), Ω (U03A9), Ϟ (U03E0)

NOTE 1 Stigma Ϛ (U03DA / U03DB), Koppa / Qoppa Ϸ (U03DE / U03DF) and Sampi Ϟ (U03E0 / U03E1) are archaic letters that are currently used to designate numerals. Digamma Ϝ (U03DC) is not used in any modern language.

NOTE 2 Through collection 9 "Greek Symbols and Coptic" of ISO/IEC 10646:2003 MES contains also a number of Coptic letters. Their order is specified in ISO/IEC 14651.

Cyrillic script:

а (U0430), А (U0410), ә (U04D1), Ă (U04D0), ä (U04D3), Ä (U04D2), ә (U04D9), Ә (U04D8),
ӓ (U04DB), Ӑ (U04DA), æ (U04D5), Æ (U04D4), б (U0431), Б (U0411), в (U0432), В (U0412),
г (U0433), Г (U0413), ґ (U0493), Ґ (U0492), ҕ (U0495), Ҙ (U0494), д (U0434), Д (U0414),
ђ (U0452), Ѓ (U0402), ӡ (U0499), Ӣ (U0498), е (U0435), Е (U0415), ё (U04D7), Ё (U04D6),
є (U0454), Є (U0404), ж (U0436), Ж (U0416), Ӈ (U04DD), Ә (U04DC), ӗ (U0497), Ә (U0496),
з (U0437), З (U0417), ӧ (U04DF), Ӛ (U04DE), s (U0455), S (U0405), ӟ (U04E1), Ӡ (U04E0),
и (U0438), И (U0418), ӡ (U04E5), Ӣ (U04E4), i (U0456), I (U0406), ı (U0457), İ (U0407),
й (U0439), Й (U0419), ӥ (U048B), Ӧ (U048A), j (U0458), J (U0408), к (U043A), К (U041A),
қ (U049B), Қ (U049A), ӕ (U04C4), Ӧ (U04C3), к (U04A1), К (U04A0), ӗ (U049F), Ә (U049E),
к (U049D), К (U049C), л (U043B), Л (U041B), л (U04C6), Ӡ (U04C5), ӡ (U0459), Ӣ (U0409),
м (U043C), М (U041C), ӡ (U04CE), Ӣ (U04CD), н (U043D), Н (U041D), ӡ (U04CA), Ӣ (U04C9),

Ң (U04A3), Ҥ (U04A2), Ҧ (U04C8), Ҩ (U04C7), Ҭ (U04A5), Ү (U04A4), Ұ (U045A), Ҳ (U040A),
 о (U043E), О (U041E), ö (U04E7), Ö (U04E6), ø (U04E9), Ø (U04E8), ø (U04EB), Æ (U04EA),
 п (U043F), П (U041F), Ӏ (U04A7), ӂ (U04A6), с (U0481), С (U0480), р (U0440), Р (U0420),
 р (U048F), Р (U048E), с (U0441), С (U0421), ç (U04AB), Ç (U04AA), т (U0442), Т (U0422),
 т (U04AD), Ӣ (U04AC), ӥ (U045B), Ӧ (U040B), у (U0443), У (U0423), ӱ (U045E), Ӳ (U040E),
 ӳ (U04F1), Ӵ (U04F0), ӵ (U04F3), Ӷ (U04F2), Ү (U04AF), Ү (U04AE), Ұ (U04B1), Ҳ (U04B0),
 ӧ (U0479), Ө (U0478), Ӧ (U0444), Ӧ (U0424), x (U0445), X (U0425), x (U04B3), Ӧ (U04B2),
 h (U04BB), H (U04BA), w (U0461), Ӧ (U0460), Ӧ (U047F), Ӧ (U047E), Ӧ (U047D), Ӧ (U047C),
 Ӧ (U047B), Ӧ (U047A), Ӧ (U0446), Ӧ (U0426), Ӧ (U04B5), Ӧ (U04B4), ч (U0447), Ч (U0427),
 Ӧ (U04F5), Ӧ (U04F4), ч (U04B7), Ч (U04B6), ч (U04CC), Ч (U04CB), ч (U04B9), Ч (U04B8),
 Ӧ (U04BD), Ӧ (U04BC), Ӧ (U04BF), Ӧ (U04BE), Ӧ (U045F), Ӧ (U040F), Ӧ (U0448), Ӧ (U0428),
 Ӧ (U0449), Ӧ (U0429), Ӧ (U044A), Ӧ (U042A), Ӧ (U044B), Ӧ (U042B), Ӧ (U04F9), Ӧ (U04F8),
 Ӧ (U044C), Ӧ (U042C), Ӧ (U048D), Ӧ (U048C), Ӧ (U0463), Ӧ (U0462), Ӧ (U044D), Ӧ (U042D),
 Ӧ (U04ED), Ӧ (U04EC), ю (U044E), Ю (U042E), я (U044F), Я (U042F), Ӧ (U0465), Ӧ (U0464),
 Ӧ (U0467), Ӧ (U0466), Ӧ (U046B), Ӧ (U046A), Ӧ (U0469), Ӧ (U0468), Ӧ (U046D), Ӧ (U046C),
 Ӧ (U046F), Ӧ (U046E), Ӧ (U0471), Ӧ (U0470), Ӧ (U0473), Ӧ (U0472), Ӧ (U0475), Ӧ (U0474),
 Ӧ (U0477), Ӧ (U0476), Ӧ (U04A9), Ӧ (U04A8), I (U04C0)

Georgian script:

ა (U10D0), ბ (U10D1), გ (U10D2), დ (U10D3), ე (U10D4), ვ (U10D5), ზ (U10D6), თ (U10F1),
 ი (U10D7), კ (U10D8), ლ (U10D9), მ (U10DA), ნ (U10DB), ე (U10DC), ვ (U10F2), თ (U10DD),
 პ (U10DE), ჟ (U10DF), რ (U10E0), ს (U10E1), ტ (U10E2), ჯ (U10F3), ლ (U10E3), მ (U10E4),
 ქ (U10E5), ლ (U10E6), გ (U10E7), მ (U10E8), ნ (U10E9), ე (U10EA), დ (U10EB), ვ (U10EC),
 ჯ (U10ED), ს (U10EE), ჯ (U10F4), ჯ (U10EF), ჯ (U10F0), მ (U10F5), ლ (U10F6), გ (U10F7),
 ე (U10F8)

NOTE 3 ȝ (U10F3), ȝ (U10F4), ȝ (U10F5) and ȝ (U10F6) are today considered archaic letters.

Armenian script:

ա (U0561), Ա (U0531), բ (U0562), Բ (U0532), գ (U0563), Գ (U0533), դ (U0564), Դ (U0534),
 ե (U0565), Ե (U0535), զ (U0566), Զ (U0536), է (U0567), Է (U0537), ը (U0568), Ը (U0538),
 փ (U0569), Փ (U0539), ժ (U056A), Ժ (U053A), ի (U056B), Ի (U053B), լ (U056C), Լ (U053C),
 յ (U056D), Կ (U053D), ծ (U056E), Ծ (U053E), կ (U056F), Կ (U053F), հ (U0570), Հ (U0540),
 ճ (U0571), Ջ (U0541), ղ (U0572), Ղ (U0542), ճ (U0573), Ճ (U0543), տ (U0574), Թ (U0544),
 յ (U0575), Յ (U0545), ն (U0576), Ն (U0546), շ (U0577), Շ (U0547), ո (U0578), Ո (U0548),
 չ (U0579), Չ (U0549), պ (U057A), Պ (U054A), ջ (U057B), Ջ (U054B), ռ (U057C), Ր (U054C),
 ս (U057D), Ս (U054D), վ (U057E), Վ (U054E), տ (U057F), Թ (U054F), ը (U0580), Ը (U0550),
 ց (U0581), Ծ (U0551), լ (U0582), Լ (U0552), փ (U0583), Փ (U0553), ք (U0584), Ք (U0554),
 ւ (U0587), Օ (U0585), Օ (U0555), ֆ (U0586), Ֆ (U0556)

A.1.4

diacritical mark

any of a number of recurring graphical structures placed over, under or next to a first level letter which does not significantly modify the shape of the first level letter itself and which in combination with that first level letter is a valid letter

NOTE These structures modify meaning or pronunciation or some other feature of the first level letter. The diacritical marks which are relevant to this European Standard are listed in A.8.1.

A.1.5

letter with diacritical marks

letter which can be seen as equivalent to the combination between a first level letter and one or more diacritical marks

NOTE Some *letters with diacritical marks* are treated as *first level letters* in some languages, e.g. ä in Swedish and ñ in Spanish. However, these are subject to national standards or local practices which are outside the scope of this European Standard.

NOTE Very few Latin letters such as á (U01FB) have more than one diacritical mark. A considerable number of Greek letters have more than one diacritical mark.

A.1.6

equivalent letter form

character created by joining two or more distinct first level letters or two or more letters with diacritical marks or any combination of these

NOTE Examples for equivalent letter forms in MES are the LATIN SMALL LIGATURE FI (UFB01), LATIN SMALL LIGATURE FL (UFB02), and the Croatian dz (U01F2 / U01F3).

A.1.7

second level letter

letter that is neither a first level letter nor an equivalent letter form nor a letter with diacritical marks

NOTE The second level letters which are relevant to this European Standard are listed in A.8.2

A.1.8

capital letter

letter which has the string CAPITAL in its name in ISO/IEC 10646

NOTE 1 This definition works for the repertoire of MES, but not necessarily for the full repertoire of the UCS.

NOTE 2 A capital letter is also known as an uppercase letter

NOTE 3 For the first level letters these are:

Latin script:

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 Ъ

Greek script:

Α Β Γ Δ Ε Ζ Η Θ Ι Κ Λ Μ Ν Ξ Ο Π Ρ Σ Τ Υ Φ Χ Ψ Ω

Cyrillic script:

А Ă Ä Ə Æ Б В Г Ƣ Д Ƣ ƣ Е Ё Є Ж Ӣ Ӥ Ӧ С Ƶ И Й І Ї Ӏ Ӑ Ј
К Ƙ ƚ ƞ Ɵ Ơ ơ Ƣ ƣ Ƥ ƥ Ʀ Ƨ ƨ Ʃ ƪ ƫ Ƭ Ʒ Ƹ ƹ ƺ ƻ Ƽ ƾ ƿ ƺ ƻ Ƽ ƾ ƿ ƺ ƻ Ƽ ƾ ƿ
Ҁ ҁ ҂ ҃ ҄ ҅ ҆ ҇ ҈ ҉ Ҋ ҋ Ҍ ҍ Ҏ ҏ Ґ ґ Ғ Ҕ Җ Ҙ Қ қ Ҝ ҝ Ҟ ҟ Ҡ ҡ Ң Ҥ ҥ Ҧ ҧ Ҩ ҩ Ҫ ҫ ҭ Ү Ұ Ҳ Ҵ Ҷ Ҹ ҹ Һ һ Ҽ ҽ Ҿ ҿ Ұ Ҳ Ҵ Ҷ Ҹ ҹ Һ һ Ҽ ҽ Ҿ ҿ
Ӏ Ӑ ӑ Ӓ ӓ Ӕ ӕ Ӗ ӗ Ә ә Ӛ ӛ Ӝ ӝ Ӟ ӟ Ӡ ӡ Ӣ ӣ Ӥ ӥ Ӧ ӧ Ө ө Ӫ ӫ Ӭ ӭ Ӯ ӯ Ӱ ӱ Ӳ ӳ Ӵ ӵ Ӷ ӷ Ӹ ӹ Ӻ ӻ Ӽ ӽ Ӿ ӿ Ӏ Ӑ ӑ Ӓ ӓ Ӕ ӕ Ӗ ӗ Ә ә Ӛ ӛ Ӝ ӝ Ӟ ӟ Ӡ ӡ Ӣ ӣ Ӥ ӥ Ӧ ӧ Ө ө Ӫ ӫ Ӭ ӭ Ӯ ӯ Ӱ ӱ Ӳ ӳ Ӵ ӵ Ӷ ӷ Ӹ ӹ Ӻ ӻ Ӽ ӽ Ӿ ӿ

Georgian script:

(none in MES)

NOTE 4 The function of capital letters in Georgian differs significantly from the function of capital letters in the other four scripts. The Georgian letters which ISO/IEC 10646 calls GEORGIAN CAPITAL LETTER make up the asomtavruli script that is primarily used in Old Georgian texts. The remaining letters (classified simply as GEORGIAN LETTER in ISO/IEC 10646) are usually identified with the mxedruli or military script that is used almost exclusively for writing modern Georgian. For this reason the MES collection only comprises the "Basic Georgian" collection (10D0-10FF) with the mxedruli script.

Armenian script:

Ա Բ Գ Դ Ե Զ Է Ը Թ Ժ Ի Լ Խ Ծ Կ Ը Թ Ճ Ծ Ն Շ Ո Զ Պ Ջ Ռ Ս Վ Շ Ի Փ
Ք օ Ֆ

A.1.9

small letter

letter which is not a capital letter

NOTE A small letter is also known as a lowercase letter

A.1.10

special character

character that is neither a letter nor a digit

NOTE *Special characters* are often called *symbols*, but also include punctuation marks, apostrophes, mathematical operators, monetary symbols and others.

A.1.11

space character

one of the special characters listed in 20.1 of ISO/IEC 10646:2003

NOTE 1 The space characters also in MES are 0020 SPACE, 00A0 NO-BREAK SPACE, 2000 EN QUAD, 2001 EM QUAD, 2002 EN SPACE, 2003 EM SPACE, 2004 THREE-PER-EM SPACE, 2005 FOUR-PER-EM SPACE, 2006 SIX-PER-EM SPACE, 2007 FIGURE SPACE, 2008 PUNCTUATION SPACE, 2009 THIN SPACE, 200A HAIR SPACE

NOTE 2 A number of different types of "spaces" such as tabulators or line breaks exist which are not part of MES, but which are used often in various fields of application. These may be understood as space characters for the purposes of this annex.

A.2 Preparatory procedures

A.2.1 Purpose

Most ordering tasks require more than simply the 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 identical.

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

A.2.2 Methodology

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:

- 1) subdivision of data into multiple ordering keys through the introduction of a higher level protocol;
- 2) establishing a hierarchy between these keys;
- 3) extracting the keys from the data;
- 4) subjecting the keys to some form of normalization.

NOTE 1 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, or reduction of blanks between words to one throughout the data. It can also be left out entirely.

NOTE 2 An especially important step is usually the correct treatment of numeral strings where leading zeroes might have to be introduced to ensure proper comparisons between corresponding decimals. Failure to do so may result in faulty ordering.

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 users have already performed these preparatory procedures which are left entirely to their 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 informative annex a multilevel ordering procedure whose results are identical to those produced by the application of the rules of the body of this European Standard.

Multilevel ordering procedure means that the input strings are first compared on the *first 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 *third ordering level* is invoked, and after this 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. If 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 according to certain assumptions:

- access to information must be facilitated as much as possible;
- the user is not assumed to know details of ISO/IEC 10646;
- 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.

These assumptions motivate a set of principles that underlie these European Ordering Rules and help to clarify the decisions taken:

- second level 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 should likewise follow after more basic forms.

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 before Georgian ones before Armenian ones.

A.3.3.2 Equivalent letter forms

Equivalent letter forms are decomposed into the letters out of which they are formed.

A.4 First ordering level

A.4.1 Validity

All of the following rules are valid for the first 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 Second level letters

Second level letters are treated as equivalent to one or more first level letters as specified in A.8.2.

A.4.2.3 Letters with diacritical marks

Letters with diacritical marks are treated as equivalent to their corresponding first level letters.

NOTE For the definition of first level letters please cf. A.1.3.

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

Latin first level letters are to be ordered in the following sequence:

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 þ

A.4.3.3 Greek script

Greek first level letters are to be ordered in the following sequence:

A.5.3 Ordering sequences

A.5.3.1 Second level letters

Second level letters are to be ordered after their corresponding first level letter. In the case of multiple second level letters with the same first level letter they are to be ordered in the sequence specified by A.8.2.

A.5.3.2 Letters with diacritical marks

Letters with diacritical marks that have only one diacritical mark are to be ordered with respect to their diacritical mark in the sequence indicated in A.8.1.1. For letters with more than one diacritical mark, the diacritical mark shall be considered in the following order: Inside the character before outside; below the character before above; working from bottom to top; then from left to right. In practice, this results for MES in the sequence indicated in A.8.1.2.

NOTE In some European countries, notably in France, diacritics are treated differently from this European Standard by parsing diacritics backwards within each word. For applications targeted for this market this should be taken into consideration by the declaration of a suitable delta.

A.6 Third ordering level

A.6.1 No unique sequence after the second ordering level

If the second ordering level also does not result in a unique sequence of strings, the third ordering level is invoked. It no longer treats 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.

A.7 Fourth ordering level

A.7.1 No unique sequence after the third ordering level

If the third 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

Most special characters of the *Multilingual European Subset No 3* except for currency signs are ordered in the sequence of the default tailorable template of ISO/IEC 14651. For most special characters this is the order in which they are listed in ISO/IEC 10646 and relevant appendices. However, for a number of special characters ISO/IEC 14651 defines a divergent sequence in line with the specification of the Canadian standard CAN/CSA Z243.230-1998.

NOTE It is advised to pay particular attention to special characters that may have the role of structuring entries in some manner. These include punctuation marks, hyphens, apostrophes and brackets.

A.7.3 Equivalence

Two strings between which after the fourth ordering level no unique sequence can be established are considered to be equivalent.

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

A.8 Specific ordering sequences

A.8.1 Diacritical marks

This form of presentation has been chosen to enable the unification of diacritical marks across scripts without modifying the resulting sequence of strings. Official Greek names of the diacritics are underlined.

Table A.1 — Diacritical marks

Shape ^a	Diacritical mark ^b		Alternative names ^c
´	U0313	COMBINING COMMA ABOVE	<u>PSILI</u> (spacing U1FBF) / spiritus lenis
ˆ	U0314	COMBINING REVERSED COMMA ABOVE	<u>DASIA</u> (spacing U1FFE) / spiritus asper
ˊ	U0301	COMBINING ACUTE ACCENT	<u>OXIA</u> , Tonos
ˋ	U0300	COMBINING GRAVE ACCENT	<u>VARIA</u>
˘	U0306	COMBINING BREVE	<u>VRACHY</u>
ˆ	U0302	COMBINING CIRCUMFLEX ACCENT	
ˇ	U030C	COMBINING CARON	
◌ ^o	U030A	COMBINING RING ABOVE	
˜	U0342	COMBINING GREEK <u>PERISPOMENI</u>	
¨	U0308	COMBINING DIAERESIS	<u>DIALYTICA</u> , umlaut, trema ^d
˝	U030B	COMBINING DOUBLE ACUTE ACCENT	
˜	U0303	COMBINING TILDE	
˙	U0307	COMBINING DOT ABOVE	
¸	U0327	COMBINING CEDILLA	
˛	U0328	COMBINING OGONEK	
ˉ	U0304	COMBINING MACRON	Greek macron, length
¸	U0326	COMBINING COMMA BELOW ^e	
ι	U1FBE	<u>PROSGEGRAMMENI</u>	iota adscriptum
ι	U0345	COMBINING GREEK <u>YPOGEGRAMMENI</u>	iota subscriptum ^f

- ^a Shapes may vary according to fonts and styles.
- ^b 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.
- ^c Names in lowercase letters are only an informative selection of some of the most common alternative names. Names in capitals are normative.
- ^d Strictly speaking, umlaut and trema can be two typographically slightly different phenomena, but the distinction is increasingly becoming obsolete.
- ^e The letters sometimes referred to as small g with comma above and capital g with comma below are to be ordered as small g with cedilla and capital g with cedilla respectively.
- ^f Exists only in combination with α, η, ω as ρ, ϱ, ϖ.

Table A.2 — Multiple diacritical marks

Shape	Diacritical mark ^a	
ǎ ^b	U1FCE	PSILI AND OXIA
ạ̌	—	PSILI AND OXIA AND YPOGEGRAMMENI
ã	U1FCD	PSILI AND VARIA
ạ̃	—	PSILI AND VARIA AND YPOGEGRAMMENI
ã̃	U1FCF	PSILI AND PERISPOMENI
ạ̃̃	—	PSILI AND PERISPOMENI AND YPOGEGRAMMENI
ã̇	—	PSILI AND YPOGEGRAMMENI
ǎ̃	U1FDE	DASIA AND OXIA
ạ̌̃	—	DASIA AND OXIA AND YPOGEGRAMMENI
ã̃	U1FDD	DASIA AND VARIA
ạ̃̃	—	DASIA AND VARIA AND YPOGEGRAMMENI
ǎ̃̃	U1FDF	DASIA AND PERISPOMENI
ạ̌̃̃	—	DASIA AND PERISPOMENI AND YPOGEGRAMMENI
ǎ̇̃	—	DASIA AND YPOGEGRAMMENI
ǎ̇̃̃	—	OXIA AND YPOGEGRAMMENI
ã̇̃	—	VARIA AND YPOGEGRAMMENI
á̇	—	RING ABOVE AND ACUTE
ã̇̃̃̃	—	PERISPOMENI AND YPOGEGRAMMENI
ˆ	U1FEE	DIALYTIKA AND OXIA
ˆ	U0385	DIALYTIKA AND TONOS

˘	U1FED	DIALYTIKA AND VARIA
̂	U1FC1	DIALYTIKA AND PERISPOMENI
̄ ö	—	DIAERESIS AND MACRON

^a Position and name in ISO/IEC 10646. 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 across scripts unless this reflects prevalent usage and user-expectations.

^b This and several other combinations cannot reasonably be printed as stand-alone diacritics. They are presented here in combination with the letter α.

A.8.2 Second level letters

Table A.3 — Second level letters

Shape	Position and name of second level letter in ISO/IEC 10646		Equiv. FOL ^a
ⱱ	U0250	LATIN SMALL LETTER TURNED A	a
α	U0251	LATIN SMALL LETTER ALPHA	α
Ɀ	U0252	LATIN SMALL LETTER TURNED ALPHA	a
æ	U00E6	LATIN SMALL LETTER AE	æ
Æ	U00C6	LATIN CAPITAL LETTER AE	Æ
ǣ	U01FD	LATIN SMALL LETTER AE WITH ACUTE	ǣ
Ǽ	U01FC	LATIN CAPITAL LETTER AE WITH ACUTE	Ǽ
ǣ̄	U01E3	LATIN SMALL LETTER AE WITH MACRON	ǣ̄
Ǽ̄	U01E2	LATIN CAPITAL LETTER AE WITH MACRON	Ǽ̄
Ɱ	U0299	LATIN LETTER SMALL CAPITAL B	b
ƀ	U0180	LATIN SMALL LETTER B WITH STROKE	b
Ɓ	U0243	LATIN CAPITAL LETTER B WITH STROKE	Ɓ
ƃ	U0253	LATIN SMALL LETTER B WITH HOOK	b
Ƅ	U0181	LATIN CAPITAL LETTER B WITH HOOK	Ƅ
Ḃ	U0183	LATIN SMALL LETTER B WITH TOPBAR	b
Ḅ	U0182	LATIN CAPITAL LETTER B WITH TOPBAR	Ḅ
ç	U0188	LATIN SMALL LETTER C WITH HOOK	c
Ɔ	U0187	LATIN CAPITAL LETTER C WITH HOOK	Ɔ
ç̣	U0255	LATIN SMALL LETTER C WITH CURL	c

ċ	U0297	LATIN LETTER STRETCHED C	ċ
đ	U0111	LATIN SMALL LETTER D WITH STROKE	đ
Ð	U0110	LATIN CAPITAL LETTER D WITH STROKE	Ð
ð	U00F0	LATIN SMALL LETTER ETH	ð
Ð	U00D0	LATIN CAPITAL LETTER ETH	Ð
ꞥ	U0256	LATIN SMALL LETTER D WITH TAIL	ꞥ
Ð	U0189	LATIN CAPITAL LETTER AFRICAN D	Ð
ɖ	U0257	LATIN SMALL LETTER D WITH HOOK	ɖ
Ɖ	U018A	LATIN CAPITAL LETTER D WITH HOOK	Ɖ
ɗ	U018C	LATIN SMALL LETTER D WITH TOPBAR	ɗ
Ɗ	U018B	LATIN CAPITAL LETTER D WITH TOPBAR	Ɗ
ɔ̣	U0221	LATIN SMALL LETTER D WITH CURL	ɔ̣
ɔ̥	U018D	LATIN SMALL LETTER TURNED DELTA	ɔ̥
ɖ͡	U02A5	LATIN SMALL LETTER DZ DIGRAPH WITH CURL	ɖ͡
ɖ͡	U02A4	LATIN SMALL LETTER DEZH DIGRAPH	ɖ͡
ə̣	U01DD	LATIN SMALL LETTER TURNED E	ə̣
Ǝ	U018E	LATIN CAPITAL LETTER REVERSED E	Ǝ
ə̥	U0259	LATIN SMALL LETTER SCHWA	ə̥
Ɛ	U018F	LATIN CAPITAL LETTER SCHWA	Ɛ
ɛ̥	U025B	LATIN SMALL LETTER OPEN E	ɛ̥
Ɛ	U0190	LATIN CAPITAL LETTER OPEN E	Ɛ
ɛ̣	U0258	LATIN SMALL LETTER REVERSED E	ɛ̣
ɛ̣̥	U025A	LATIN SMALL LETTER SCHWA WITH HOOK	ɛ̣̥
ɛ̣̥	U025C	LATIN SMALL LETTER REVERSED OPEN E	ɛ̣̥
ɛ̣̥̥	U025D	LATIN SMALL LETTER REVERSED OPEN E WITH HOOK	ɛ̣̥̥
ɛ̣̥̥̥	U025E	LATIN SMALL LETTER CLOSED REVERSED OPEN E	ɛ̣̥̥̥
ɛ̣̥̥̥̥	U029A	LATIN SMALL LETTER CLOSED OPEN E	ɛ̣̥̥̥̥
ƒ	U0192	LATIN SMALL LETTER F WITH HOOK	ƒ
Ƒ	U0191	LATIN CAPITAL LETTER F WITH HOOK	Ƒ

g	U0261	LATIN SMALL LETTER SCRIPT G	g
Ɔ	U0262	LATIN LETTER SMALL CAPITAL G	g
g̃	U01E5	LATIN SMALL LETTER G WITH STROKE	g
Ɔ̃	U01E4	LATIN CAPITAL LETTER G WITH STROKE	G
g̣	U0260	LATIN SMALL LETTER G WITH HOOK	g
Ɔ̣	U0193	LATIN CAPITAL LETTER G WITH HOOK	G
g̤	U029B	LATIN SMALL CAPITAL LETTER G WITH HOOK	g
γ	U0263	LATIN SMALL LETTER GAMMA	g
Υ	U0194	LATIN CAPITAL LETTER GAMMA	G
ʁ	U0264	LATIN SMALL LETTER RAMS HORN	g
ɔ ^b	U01A3	LATIN SMALL LETTER OI	g
Ɔ	U01A2	LATIN CAPITAL LETTER OI	G
ħ	U0127	LATIN SMALL LETTER H WITH STROKE	h
Ĥ	U0126	LATIN CAPITAL LETTER H WITH STROKE	H
Ɔ	U029C	LATIN LETTER SMALL CAPITAL H	h
h̃	U0266	LATIN SMALL LETTER H WITH HOOK	h
ḥ	U0267	LATIN SMALL LETTER HENG WITH HOOK	h
ɥ	U0265	LATIN SMALL LETTER TURNED H	h
ɥ̣	U02AE	LATIN SMALL LETTER TURNED H WITH FISHHOOK	h
ɥ̣̅	U02AF	LATIN SMALL LETTER TURNED H WITH FISHHOOK AND TAIL	h
h̥	U0195	LATIN SMALL LETTER HV	hv
Ĥ	U01F6	LATIN CAPITAL LETTER HWAIR	HV
ı	U2071	SUPERSCRIPT LATIN SMALL LETTER I	i
ɿ	U0131	LATIN SMALL LETTER DOTLESS I	i
Ɔ	U026A	LATIN LETTER SMALL CAPITAL I	i
ı̃	U0268	LATIN SMALL LETTER I WITH STROKE	i
Ɔ̃	U0197	LATIN CAPITAL LETTER I WITH STROKE	I
ι	U0269	LATIN SMALL LETTER IOTA	i
Ɔ	U0196	LATIN CAPITAL LETTER IOTA	I

ij	U0133	LATIN SMALL LIGATURE IJ	ij
IJ	U0132	LATIN CAPITAL LIGATURE IJ	IJ
ĵ	U029D	LATIN SMALL LETTER J WITH CROSSED-TAIL	ĵ
ƚ	U025F	LATIN SMALL LETTER DOTLESS J WITH STROKE	ƚ
ƚ	U0284	LATIN SMALL LETTER DOTLESS J WITH STROKE AND HOOK	ƚ
ƙ	U0199	LATIN SMALL LETTER K WITH HOOK	ƙ
Ƙ	U0198	LATIN CAPITAL LETTER K WITH HOOK	Ƙ
κ	U0138	LATIN SMALL LETTER KRA	κ
ƚ	U029E	LATIN SMALL LETTER TURNED K	ƚ
ł	U0142	LATIN SMALL LETTER L WITH STROKE	ł
Ł	U0141	LATIN CAPITAL LETTER L WITH STROKE	Ł
ł̇	U0140	LATIN SMALL LETTER L WITH MIDDLE DOT	ł̇
Ł̇	U013F	LATIN CAPITAL LETTER L WITH MIDDLE DOT	Ł̇
Ɑ	U029F	LATIN LETTER SMALL CAPITAL L	Ɑ
ł̄	U019A	LATIN SMALL LETTER L WITH BAR	ł̄
Ł̄	U023D	LATIN CAPITAL LETTER L WITH BAR	Ł̄
ł̃	U026B	LATIN SMALL LETTER L WITH MIDDLE TILDE	ł̃
ł̣	U026C	LATIN SMALL LETTER L WITH BELT	ł̣
ł̡	U026D	LATIN SMALL LETTER L WITH RETROFLEX HOOK	ł̡
ł̢	U0234	LATIN SMALL LETTER L WITH CURL	ł̢
λ̣	U019B	LATIN SMALL LETTER LAMBDA WITH STROKE	λ̣
ł̣	U026E	LATIN SMALL LETTER LEZH	ł̣
ɱ	U0271	LATIN SMALL LETTER M WITH HOOK	ɱ
ɯ	U026F	LATIN SMALL LETTER TURNED M	ɯ
Ɲ	U019C	LATIN CAPITAL LETTER TURNED M	Ɲ
ɱ	U0270	LATIN SMALL LETTER TURNED M WITH LONG LEG	ɱ
ⁿ	U207F	SUPERSCRIPT LATIN SMALL LETTER N	ⁿ
'n	U0149	LATIN SMALL LETTER N PRECEDED BY APOSTROPHE	'n
Ɱ	U0274	LATIN LETTER SMALL CAPITAL N	Ɱ

ɲ	U0272	LATIN SMALL LETTER N WITH LEFT HOOK	ɲ
Ń	U019D	LATIN CAPITAL LETTER N WITH LEFT HOOK	Ń
ɳ	U019E	LATIN SMALL LETTER N WITH LONG RIGHT LEG	ɳ
Ť	U0220	LATIN CAPITAL LETTER N WITH LONG RIGHT LEG	Ť
ɳ̥	U0273	LATIN SMALL LETTER N WITH RETROFLEX HOOK	ɳ̥
ɳ̂	U0235	LATIN SMALL LETTER N WITH CURL	ɳ̂
ɳ̌	U014B	LATIN SMALL LETTER ENG	ɳ̌
Ŋ	U014A	LATIN CAPITAL LETTER ENG	Ŋ
ø	U00F8	LATIN SMALL LETTER O WITH STROKE	ø
Ø	U00D8	LATIN CAPITAL LETTER O WITH STROKE	Ø
ø̇	U01FF	LATIN SMALL LETTER O WITH STROKE AND ACUTE	ø̇
Ø̇	U01FE	LATIN CAPITAL LETTER O WITH STROKE AND ACUTE	Ø̇
ɔ̥	U01A1	LATIN SMALL LETTER O WITH HORN	ɔ̥
Ɔ	U01A0	LATIN CAPITAL LETTER O WITH HORN	Ɔ
ɔ̜	U0254	LATIN SMALL LETTER OPEN O	ɔ̜
Ɔ̜	U0186	LATIN CAPITAL LETTER OPEN O	Ɔ̜
ø̄	U0275	LATIN SMALL LETTER BARRED O	ø̄
Ө	U019F	LATIN CAPITAL LETTER O WITH MIDDLE TILDE	Ө
ω	U0277	LATIN SMALL LETTER CLOSED OMEGA	ω
ø̧	U0223	LATIN SMALL LETTER OU	ø̧
Ɔ̧	U0222	LATIN CAPITAL LETTER OU	Ɔ̧
œ	U0153	LATIN SMALL LIGATURE OE	œ
Œ	U0276	LATIN LETTER SMALL CAPITAL OE	Œ
Ɔ̥	U0152	LATIN CAPITAL LIGATURE OE	Ɔ̥
ɸ	U01A5	LATIN SMALL LETTER P WITH HOOK	ɸ
Ɔ	U01A4	LATIN CAPITAL LETTER P WITH HOOK	Ɔ
ϕ	U0278	LATIN SMALL LETTER PHI	ϕ
ɸ̇	U02A0	LATIN SMALL LETTER Q WITH HOOK	ɸ̇
Ŕ	U0280	LATIN LETTER SMALL CAPITAL R	Ŕ

Ŕ	U01A6	LATIN LETTER YR	R
ɹ	U0279	LATIN SMALL LETTER TURNED R	r
ɹ̇	U027A	LATIN SMALL LETTER TURNED R WITH LONG LEG	r
ɹ̣	U027B	LATIN SMALL LETTER TURNED R WITH HOOK	r
ɹ̥	U027C	LATIN SMALL LETTER R WITH LONG LEG	r
ɹ̦	U027D	LATIN SMALL LETTER R WITH TAIL	r
ɹ̧	U027E	LATIN SMALL LETTER R WITH FISHHOOK	r
ɹ̨	U027F	LATIN SMALL LETTER REVERSED R WITH FISHHOOK	r
ʀ	U0281	LATIN LETTER SMALL CAPITAL INVERTED R	r
ſ	U017F	LATIN SMALL LETTER LONG S	s
ſ̃	U0282	LATIN SMALL LETTER S WITH HOOK	s
ſ̥	U0283	LATIN SMALL LETTER ESH	s
Σ	U01A9	LATIN CAPITAL LETTER ESH	S
ʒ	U01AA	LATIN REVERSED ESH LOOP	s
ʒ̣	U0285	LATIN SMALL LETTER SQUAT REVERSED ESH	s
ſ̧	U0286	LATIN SMALL LETTER ESH WITH CURL	s
ß	U00DF	LATIN SMALL LETTER SHARP S	ss
ƚ	U0167	LATIN SMALL LETTER T WITH STROKE	t
Ɖ	U0166	LATIN CAPITAL LETTER T WITH STROKE	T
ƚ̣	U01AB	LATIN SMALL LETTER T WITH PALATAL HOOK	t
ƚ̥	U01AD	LATIN SMALL LETTER T WITH HOOK	t
Ɖ̥	U01AC	LATIN CAPITAL LETTER T WITH HOOK	T
ƚ̦	U0288	LATIN SMALL LETTER T WITH RETROFLEX HOOK	t
Ɖ̦	U01AE	LATIN CAPITAL LETTER T WITH RETROFLEX HOOK	T
ƚ̧	U0236	LATIN SMALL LETTER T WITH CURL	t
ɹ̥	U0287	LATIN SMALL LETTER TURNED T	t
ts̥	U02A8	LATIN SMALL LETTER TC DIGRAPH WITH CURL	tc
ụ	U01B0	LATIN SMALL LETTER U WITH HORN	u
Ū	U01AF	LATIN CAPITAL LETTER U WITH HORN	U

Ɑ	U0289	LATIN SMALL LETTER U BAR	u
Ɱ	U0244	LATIN CAPITAL LETTER U BAR	U
υ	U028A	LATIN SMALL LETTER UPSILON	u
Ϝ	U01B1	LATIN CAPITAL LETTER UPSILON	U
ϵ	U028B	LATIN SMALL LETTER V WITH HOOK	v
Ϟ	U01B2	LATIN CAPITAL LETTER V WITH HOOK	V
Ϸ	U028C	LATIN SMALL LETTER TURNED V	v
ϸ	U0245	LATIN CAPITAL LETTER TURNED V	V
Ϲ	U028D	LATIN SMALL LETTER TURNED W	w
Ƶ	U01BF	LATIN LETTER WYNN	w
ƶ	U01F7	LATIN CAPITAL LETTER WYNN	W
Ʒ	U028F	LATIN LETTER SMALL CAPITAL Y	y
Ƹ	U01B4	LATIN SMALL LETTER Y WITH HOOK	y
ƹ	U01B3	LATIN CAPITAL LETTER Y WITH HOOK	Y
ƺ	U028E	LATIN SMALL LETTER TURNED Y	y
ƻ	U021D	LATIN SMALL LETTER YOGH	y
Ƽ	U021C	LATIN CAPITAL LETTER YOGH	Y
ƽ	U01B6	LATIN SMALL LETTER Z WITH STROKE	z
ƿ	U01B5	LATIN CAPITAL LETTER Z WITH STROKE	Z
ƾ	U0225	LATIN SMALL LETTER Z WITH HOOK	z
ƿ	U0224	LATIN CAPITAL LETTER Z WITH HOOK	Z
ƿ	U0290	LATIN SMALL LETTER Z WITH RETROFLEX HOOK	z
ƿ	U0291	LATIN SMALL LETTER Z WITH CURL	z
ƿ	U0292	LATIN SMALL LETTER EZH	z
ƿ	U01B7	LATIN CAPITAL LETTER EZH	Z
ƿ	U01EF	LATIN SMALL LETTER EZH WITH CARON	z
ƿ	U01EE	LATIN CAPITAL LETTER EZH WITH CARON	Z
ƿ	U01B9	LATIN SMALL LETTER EZH REVERSED	z
ƿ	U01B8	LATIN CAPITAL LETTER EZH REVERSED	Z

ž	U01BA	LATIN SMALL LETTER EZH WITH TAIL	z
ž	U0293	LATIN SMALL LETTER EZH WITH CURL	z
ς	U03C2	GREEK SMALL LETTER FINAL SIGMA	σ
г	U0491	CYRILLIC SMALL LETTER GHE UPTURN	г
Ґ	U0490	CYRILLIC CAPITAL LETTER GHE UPTURN	Ґ
ѓ	U0453	CYRILLIC SMALL LETTER GJE	ѓ
Ѓ	U0403	CYRILLIC CAPITAL LETTER GJE	Ѓ
ќ	U045C	CYRILLIC SMALL LETTER KJE	ќ
Ќ	U040C	CYRILLIC CAPITAL LETTER KJE	Ќ
<p>^a Equivalent on First Ordering Level.</p> <p>^b Also known as letter gha.</p>			

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.

NOTE For the purpose of this annex, a space character can include all characters which are usually considered to divide words. Typical examples of these might be hyphens, apostrophes and brackets. Cf. also note to A.1.11.

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 separators. 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 versus letter-by-letter ordering

Letter-by-letter ordering	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 only few second level 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* subclause (A.3.3.1) space 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

C.1 Background

In some cases it is desirable to differentiate further on the *third ordering level*, e.g. in the case where definitions and ordinary usage of a word are distinguished solely by the application of some form of internal tagging. This tagging usually takes in print the form of a formatting style. Especially in lexicography it is also often thought to be desirable to distinguish 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 highlighting it with certain typographic features, e.g. italic typeface, that serves to indicate some property of the word.

C.2 Recommended rules

In line with ISO 12199 this European Standard recommends that, if the implementer deems it necessary to make this differentiation, she or he modify (A.9.2.1) (Capitalization) on the *third 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:

- a) roman abcde;
- b) boldface **abcde**;
- c) italic *abcde*;
- d) boldface italic ***abcde***;
- e) others.

Annex D (informative)

Mixed-script ordering with one predominant script

D.1 Background

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.

D.2 Suggested steps

This may involve 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;
- 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.

D.3 Explicit transliteration

A different, likewise common method is the method of explicit transliteration which selects the transliterated word such as logos and adds the original rendering in brackets.

Annex E (informative)

Defining National Deltas based on the EOR

E.1 Background

Ordering rules for European languages can benefit from unambiguous, ideally machine-processable specifications, both in the case of formal national standards and de facto standards issued by, e.g., the relevant language institutes. This work can be facilitated by basing this specification on a tailoring of the EOR. This annex gives an overview over possible approaches for writing such a delta.

E.2 Structured specification

In most cases a delta will start with a structured, but not machine-processable description of the linguistic and cultural ordering preferences. Such a specification can list sequences of letters that are distinguished on the first ordering level. For example, Norwegian could list 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 æ ø å, whereas the Polish first level letters would be a ą b c ć d e ę f g h i j k l ł m n ó o ó p r s ś t u w y z ź ż.

A structured specification can then discuss typical distinctions that are made on the second ordering level, such as ss versus ß in German or v versus w in Icelandic dictionaries. If required, it can then look at the treatment of otherwise identical words beginning with lowercase versus uppercase letters and specify a preference for the third ordering level. Should there be established procedures for the treatment of symbols, those can again be described.

In a second step, the differences between these rules and the EOR can be captured, again in words. Such rules could be "make w sort equal to v on the first ordering level, but distinguish them on the second, where w follows v" or "order æ ø å after z on the first ordering level" in Norwegian. In some cases, specific decisions in the EOR will already meet the requirements, so the delta can be small. Likewise, should there be no established cultural preferences, the delta can just implicitly rely on the EOR defaults. In line with ISO/IEC 14651 this European Standard recommends that the defined delta be as small as possible while still expressing the cultural and linguistic ordering preferences comprehensively.

This type of structured, but not machine-processable description meets the requirements of ISO/IEC 14651:2007, 6.4. In many cases, a profile at this level of structured, human-readable specification will suffice. If the rules are generally available, they can be translated into their machine-processable equivalents.

E.3 Machine-Readable Specification

Machine-processable specifications offer the additional benefit that they can be directly plugged into an operating system's locale data. All applications that build on the operating system's API will thus automatically profit from this data and order lists correctly according to the user's chosen cultural preferences.

ISO/IEC 14651:2007, 6.4 prescribes that a machine-processable delta must contain at least one valid `order_start` entry, a specification of the number of levels for comparison, a definition of symbol definition weights and a list of character definitions. However, it leaves the question of the concrete syntax to express these rules explicitly open. This European Standard recommends to use either LDML or the POSIX-oriented syntax employed in ISO/IEC 14651's CTT or both to express the delta.

It is out of the scope of this European Standard to describe the exact syntax and semantics of LDML or the syntax rules for ISO/IEC 14651's CTT. Clause 6 and Annex G can serve as practical examples of the design and implementation of two representations of a given delta. More tutorial information as well as pointers further examples and to syntax validators for both syntaxes can be found on <http://purl.oclc.org/NET/CDFG/EN13710>.

E.4 Example 1: National Delta for German

E.4.1 Structured Specification

E.4.1.1 First Ordering Level

- The first level letters are: abcdefghijklmnopqrstuvwxyz
- Digits follow letters
- Spaces are significant
- The letter thorn (þ U00FE / ð U00DE) is ordered as th
- The letters wynn (ƿ U01BF/ ƿU01F7) and ezh (Ʒ U0292 / ʒ U01B7) are ordered as y

E.4.1.2 Second Ordering Level

The Umlaut (diaeresis in ISO/IEC 10646) is the diacritic that must be taken into account before all others. The umlaut is treated as distinct from the trema and can only occur in combination with the base letters a, o and u. The sequence of the other diacritics is compatible with the EOR delta.

E.4.1.3 Third Ordering Level

Lowercase letters precede uppercase ones.

E.4.1.4 Fourth Ordering Level

No specific rules.

E.4.2 Delta against EOR in 14651-syntax

```
% -*- coding: utf-8; -*-  
% ISO/IEC 14651-conformant delta for DIN 5007-1:2004 (sample)  
  
reorder-after <BASE> %The umlaut is the diacritic with the highest priority  
collating-symbol <UMLAUT> %specifically for the weight of the umlaut (distinct from the Trema in DIN 5007)  
<UMLAUT>  
reorder-end  
  
%Digits must follow letters  
  
%For the treatment of Roman numerals as well as digits as numbers we  
%need preprocessing that is out of the scope of this profile  
  
reorder-after <TFFFF> %After the Han ideographs, but before <SFFFF>  
<S0030> % 0  
<S0031> % 1  
<S0032> % 2  
<S0033> % 3  
<S0034> % 4  
<S0035> % 5  
<S0036> % 6
```



```

<S0037> % 7
<S0038> % 8
<S0039> % 9
reorder-end

reorder-after <SFFFF>

order_start forward:forward:forward:forward

%Whitespace precedes according to 6.2.1 all other characters on the
%first level. They thus get a non-ignorable weight

<U0009> <S0009>;<BASE>;<MIN>;<U0009> % HORIZONTAL TABULATION (in 6429)
<U000A> <S000A>;<BASE>;<MIN>;<U000A> % LINE FEED (in 6429)
<U000B> <S000B>;<BASE>;<MIN>;<U000B> % VERTICAL TABULATION (in 6429)
<U000C> <S000C>;<BASE>;<MIN>;<U000C> % FORM FEED (in 6429)
<U000D> <S000D>;<BASE>;<MIN>;<U000D> % CARRIAGE RETURN (in 6429)
<U0020> <S0020>;<BASE>;<MIN>;<U0020> % SPACE

%Ligatures and the sharp S are already treated in the EOR delta in the sense of DIN 5007

<U00E4> <S0061>;<BASE><UMLAUT>;<MIN><MIN>;<U00E4> % LATIN SMALL LETTER A WITH DIAERESIS
<U00C4> <S0061>;<BASE><UMLAUT>;<CAP><MIN>;<U00C4> % LATIN CAPITAL LETTER A WITH DIAERESIS
<U00F6> <S006F>;<BASE><UMLAUT>;<MIN><MIN>;<U00F6> % LATIN SMALL LETTER O WITH DIAERESIS
<U00D6> <S006F>;<BASE><UMLAUT>;<CAP><MIN>;<U00D6> % LATIN CAPITAL LETTER O WITH DIAERESIS
<U00FC> <S0075>;<BASE><UMLAUT>;<MIN><MIN>;<U00FC> % LATIN SMALL LETTER U WITH DIAERESIS
<U00DC> <S0075>;<BASE><UMLAUT>;<CAP><MIN>;<U00DC> % LATIN CAPITAL LETTER U WITH DIAERESIS

%If we want an ordering according to 6.1.1.4.2, where Umlaute become
%base letter + e, these lines replace the previous six weight assignments
%<U00E4> "<S0061><S0065>;<BASE><BASE><UMLAUT>;<MIN><MIN><MIN>;<U00E4> % LATIN SMALL LETTER A WITH
DIAERESIS
%<U00C4> "<S0061><S0065>;<BASE><BASE><UMLAUT>;<CAP><MIN><MIN>;<U00C4> % LATIN CAPITAL LETTER A WITH
DIAERESIS
%<U00F6> "<S006F><S0065>;<BASE><BASE><UMLAUT>;<MIN><MIN><MIN>;<U00F6> % LATIN SMALL LETTER O WITH
DIAERESIS
%<U00D6> "<S006F><S0065>;<BASE><BASE><UMLAUT>;<CAP><MIN><MIN>;<U00D6> % LATIN CAPITAL LETTER O WITH
DIAERESIS
%<U00FC> "<S0075><S0065>;<BASE><BASE><UMLAUT>;<MIN><MIN><MIN>;<U00FC> % LATIN SMALL LETTER U WITH
DIAERESIS
%<U00DC> "<S0075><S0065>;<BASE><BASE><UMLAUT>;<CAP><MIN><MIN>;<U00DC> % LATIN CAPITAL LETTER U WITH
DIAERESIS

%Treatment of other Latin characters according to DIN 31638, 7.3.4.1
<U00FE> "<S0074><S0068>;<BASE><BASE><VRNT1>;<MIN><MIN><MIN>;<U00FE> % LATIN SMALL LETTER THORN (as th on the
first level)
<U00DE> "<S0074><S0068>;<BASE><BASE><VRNT1>;<CAP><CAP><MIN>;<U00DE> % LATIN CAPITAL LETTER THORN

<U01BF> <S0079>;<BASE><VRNT3>;<MIN><MIN>;<U01BF> % LATIN LETTER WYNN (as y on the first level)
<U01F7> <S0079>;<BASE><VRNT3>;<CAP><MIN>;<U01F7> % LATIN CAPITAL LETTER WYNN
<U0292> <S0079>;<BASE><VRNT4>;<MIN><MIN>;<U0292> % LATIN SMALL LETTER EZH (as y on the first level, yogh and ezh
are equated in ISO/IEC 10646)

```

```
<U01B7> <S0079>";<BASE><VRNT4>";<CAP><MIN>";<U01B7> % LATIN CAPITAL LETTER EZH  
reorder-end
```

E.5 Example 2: National Delta for Norwegian

E.5.1 Structured Specification

E.5.1.1 First Ordering Level

- The first level letters are: abcdefghijklmnopqrstuvwxyzæøå
- The letter ä is ordered as æ
- The letter ö is ordered as ø
- The letter ü is ordered as y
- The letter o with double acute is ordered as ø
- The letter u with double acute is ordered as y
- The letter þ is ordered as th

E.5.1.2 Second Ordering Level

No specific rules.

E.5.1.3 Third Ordering Level

Lowercase letters precede uppercase ones.

E.5.1.4 Fourth Ordering Level

No specific rules.

E.5.2 Delta against EOR in 14651-syntax

```
% ISO/IEC 14651-conformant delta for Norwegian (sample)  
reorder-after <S00FE>  
collating-symbol <S00E6> % LATIN SMALL LETTER AE  
collating-symbol <S00F8> % LATIN SMALL LETTER O WITH STROKE  
collating-symbol <S00E5> % LATIN SMALL LETTER A WITH RING ABOVE  
<S00E6>  
<S00F8>  
<S00E5>  
reorder-end  
  
reorder-after <SFFFF>  
  
order_start forward;forward;forward;forward  
%first level letters are abcdefghijklmnopqrstuvwxyzæøå  
%a-z as in EOR
```

%ü and ŷ as y

<U00FC> <S0079>;"<BASE><VRNT2>";"<MIN><MIN>";<U00FC> % LATIN SMALL LETTER U WITH DIAERESIS
<U00DC> <S0079>;"<BASE><VRNT2>";"<CAP><MIN>";<U00DC> % LATIN CAPITAL LETTER U WITH DIAERESIS
<U0171> <S0079>;"<BASE><VRNT3>";"<MIN><MIN>";<U0171> % LATIN SMALL LETTER U WITH DOUBLE ACUTE
<U0170> <S0079>;"<BASE><VRNT3>";"<CAP><MIN>";<U0170> % LATIN CAPITAL LETTER U WITH DOUBLE ACUTE

<U00E6> <S00E6>;<BASE>;<MIN>;<U00E6> % LATIN SMALL LETTER AE
<U00C6> <S00E6>;<BASE>;<CAP>;<U00C6> % LATIN CAPITAL LETTER AE

%ä as æ

<U00E4> <S00E6>;"<BASE><VRNT1>";"<MIN><MIN>";<U00E4> % LATIN SMALL LETTER A WITH DIAERESIS
<U00C4> <S00E6>;"<BASE><VRNT1>";"<CAP><MIN>";<U00C4> % LATIN CAPITAL LETTER A WITH DIAERESIS

<U00F8> <S00F8>;<BASE>;<MIN>;<U00F8> % LATIN SMALL LETTER O WITH STROKE
<U00D8> <S00F8>;<BASE>;<CAP>;<U00D8> % LATIN CAPITAL LETTER O WITH STROKE

%ö as ø

<U00F6> <S00F8>;"<BASE><VRNT1>";"<MIN><MIN>";<U00F6> % LATIN SMALL LETTER O WITH DIAERESIS
<U00D6> <S00F8>;"<BASE><VRNT1>";"<CAP><MIN>";<U00D6> % LATIN CAPITAL LETTER O WITH DIAERESIS

%õ also as ø

<U0151> <S00F8>;"<BASE><VRNT2>";"<MIN><MIN>";<U0151> % LATIN SMALL LETTER O WITH DOUBLE ACUTE
<U0150> <S00F8>;"<BASE><VRNT2>";"<CAP><MIN>";<U0150> % LATIN CAPITAL LETTER O WITH DOUBLE ACUTE

<U00FE> "<S0074><S0068>";"<BASE><BASE><VRNT1>";"<MIN><MIN><MIN>";<U00FE> % LATIN SMALL LETTER THORN (as th on the first level)

<U00DE> "<S0074><S0068>";"<BASE><BASE><VRNT1>";"<CAP><CAP><MIN>";<U00DE> % LATIN CAPITAL LETTER THORN

<U00E5> <S00E5>;<BASE>;<MIN>;<U00E5> % LATIN SMALL LETTER A WITH RING ABOVE
<U00C5> <S00E5>;<BASE>;<CAP>;<U00C5> % LATIN CAPITAL LETTER A WITH RING ABOVE

recorder-end

Annex F (informative)

Modern European Scripts / MES

This annex reproduces for ease of reference the definition of the Modern European Scripts collection in ISO/IEC 10646:2003, A.4.3:

Table F.1 — Modern European Scripts / MES

1	BASIC LATIN	0020-007E
2	LATIN-1 SUPPLEMENT	00A0-00FF
3	LATIN EXTENDED-A	0100-017F
4	LATIN EXTENDED-B	0180-024F
5	IPA EXTENSIONS	0250-02AF
6	SPACING MODIFIER LETTERS	02B0-02FF
7	COMBINING DIACRITICAL MARKS	0300-036F
8	BASIC GREEK	0370-03CF
9	GREEK SYMBOLS AND COPTIC	03D0-03FF
10	CYRILLIC	0400-04FF
11	ARMENIAN	0530-058F
27	BASIC GEORGIAN	10D0-10FF
30	LATIN EXTENDED ADDITIONAL	1E00-1EFF
31	GREEK EXTENDED	1F00-1FFF
32	GENERAL PUNCTUATION	2000-206F
33	SUPERSCRIPTS AND SUBSCRIPTS	2070-209F
34	CURRENCY SYMBOLS	20A0-20CF
35	COMBINING DIACRITICAL MARKS FOR SYMBOLS	20D0-20FF
36	LETTERLIKE SYMBOLS	2100-214F
37	NUMBER FORMS	2150-218F
38	ARROWS	2190-21FF
39	MATHEMATICAL OPERATORS	2200-22FF
40	MISCELLANEOUS TECHNICAL	2300-23FF
42	OPTICAL CHARACTER RECOGNITION	2440-245F
44	BOX DRAWING	2500-257F
45	BLOCK ELEMENTS	2580-259F

46	GEOMETRIC SHAPES	25A0-25FF
47	MISCELLANEOUS SYMBOLS	2600-26FF
65	COMBINING HALF MARKS	FE20-FE2F
70	SPECIALS	FFF0-FFFF
92	CYRILLIC SUPPLEMENT	0500-052F
104	LTR ALPHABETIC PRESENTATION FORMS	FB00-FB1C

Annex G (informative)

EOR Delta in LDML Syntax

```
<?xml version="1.0" encoding="utf-8"?>
<ldml>
  <identity>
    <!--
      Authors: Marc Wilhelm Küster (EN 13710 editor) and Åke Persson

      LDML version of the rules specified in EN 13710:2011, Clause 6
    -->
    <version number="1.0" cldrVersion="1.8.1"/>
    <generation date="2010-05-03"/>
    <language type="EOR"/>
  </identity>
  <collations validSubLocales="All European locales">
    <collation type="standard">
      <settings caseLevel="on" caseFirst="lower" strength="quaternary"/>
      <rules>
        <!--
          currency signs and modifier letters are ignored in EOR
        -->
        <reset><last_tertiary_ignorable/></reset>
        <i>$</i><!-- U0024 -->
        <i>¢</i><!-- U00A2 -->
        <i>£</i><!-- U00A3 -->
        <i>¤</i><!-- U00A4 -->
        <i>¥</i><!-- U00A5 -->
        <i>€</i><!-- U20A0 -->
        <i>₤</i><!-- U20A1 -->
        <i>₤</i><!-- U20A2 -->
        <i>₣</i><!-- U20A3 -->
        <i>₤</i><!-- U20A4 -->
        <i>₥</i><!-- U20A5 -->
        <i>₦</i><!-- U20A6 -->
        <i>₧</i><!-- U20A7 -->
        <i>₨</i><!-- U20A8 -->
        <i>₩</i><!-- U20A9 -->
        <i>₪</i><!-- U20AA -->
        <i>₫</i><!-- U20AB -->
        <i>€</i><!-- U20AC -->
        <i>₰</i><!-- U20AD -->
        <i>₱</i><!-- U20AE -->
        <i>₲</i><!-- U20AF -->
        <i>₳</i><!-- U20B0 -->
        <i>₴</i><!-- U20B1 -->
        <i>₵</i><!-- U20B2 -->
        <i>₶</i><!-- U20B3 -->
```

<i>ž </i><!-- U20B4 -->
<i>č </i><!-- U20B5 -->

<!-- General category Lm -->

<i>h </i><!-- U02B0 -->
<i>fi </i><!-- U02B1 -->
<i>j </i><!-- U02B2 -->
<i>r </i><!-- U02B3 -->
<i>t </i><!-- U02B4 -->
<i>l </i><!-- U02B5 -->
<i># </i><!-- U02B6 -->
<i>w </i><!-- U02B7 -->
<i>y </i><!-- U02B8 -->
<i>f </i><!-- U02BB -->
<i>g </i><!-- U02BC -->
<i>g </i><!-- U02BD -->
<i>v </i><!-- U02BE -->
<i>v </i><!-- U02BF -->
<i>g </i><!-- U02C0 -->
<i>f </i><!-- U02C1 -->
<i>x </i><!-- U02D0 -->
<i>v </i><!-- U02D1 -->
<i>y </i><!-- U02E0 -->
<i>l </i><!-- U02E1 -->
<i>s </i><!-- U02E2 -->
<i>x </i><!-- U02E3 -->
<i>f </i><!-- U02E4 -->
<i>v </i><!-- U02EE -->

<!-- Glottal stops -->

<i>ʔ </i><!-- U0294 -->
<i>ʕ </i><!-- U0295 -->
<i>Ɂ </i><!-- U0296 -->
<i>ʁ </i><!-- U0298 -->
<i>ʀ </i><!-- U02A1 -->
<i>ʁ </i><!-- U02A2 -->

<!--Latin letters: only a-z + þ are basic letters -->

<reset>a</reset> <!-- turned a, latin alpha, turned alpha -->

<s>e </s><!-- U0250 -->
<s>a </s><!-- U0251 -->
<s>ð </s><!-- U0252 -->

<reset>b</reset> <!-- small capital b, b with stroke, b with hook, b with topbar -->

<s>B </s><!-- U0299 -->
<s>b </s><!-- U0180 -->
<t>B </t><!-- U0243 -->
<s>B </s><!-- U0253 -->
<t>B </t><!-- U0181 -->
<s>B </s><!-- U0183 -->
<t>B </t><!-- U0182 -->

```
<reset>c</reset> <!-- c with hook, c with curl, stretched c -->
<s>ċ </s><!-- U0188 -->
<t>Ĉ </t><!-- U0187 -->
<s>ċ̄ </s><!-- U0255 -->
<s>ċ̇ </s><!-- U0297 -->

<reset>d</reset> <!-- African d, d with hook,
d with topbar, d with curl, turned delta -->
<s>ḏ </s><!-- U0256 -->
<t>Ḑ </t><!-- U0189 -->
<s>ḏ̄ </s><!-- U0257 -->
<t>Ḑ̄ </t><!-- U018A -->
<s>ḏ̇ </s><!-- U018C -->
<t>Ḑ̇ </t><!-- U018B -->
<s>ḏ̂ </s><!-- U0221 -->
<s>ḑ </s><!-- U018D -->

<reset>e</reset> <!-- turned e, schwa, open e, reversed e, schwa with hook,
reversed open e, reversed open e with hook, closed reversed open e, closed open e -->
<s>ə̃ </s><!-- U01DD -->
<t>Ḑ̄ </t><!-- U018E -->
<s>ə̄ </s><!-- U0259 -->
<t>Ḑ̄̄ </t><!-- U018F -->
<s>ē̄ </s><!-- U025B -->
<t>Ḑ̄̄̄ </t><!-- U0190 -->
<s>ə̄̄ </s><!-- U0258 -->
<s>ə̄̄̄ </s><!-- U025A -->
<s>ə̄̄̄̄ </s><!-- U025C -->
<s>ə̄̄̄̄̄ </s><!-- U025D -->
<s>ə̄̄̄̄̄̄ </s><!-- U025E -->
<s>ə̄̄̄̄̄̄̄ </s><!-- U029A -->

<reset>f</reset> <!-- f with hook -->
<s>ḑ̄ </s><!-- U0192 -->
<t>ḑ̄̄ </t><!-- U0191 -->

<reset>g</reset> <!-- script g, small capital g, g with stroke,
g with hook, latin gamma, gha -->
<s>ḡ </s><!-- U0261 -->
<s>ḡ̄ </s><!-- U0262 -->
<s>ḡ̇ </s><!-- U01E5 -->
<t>Ḡ </t><!-- U01E4 -->
<s>ḡ̄̄ </s><!-- U0260 -->
<t>Ḡ̄̄ </t><!-- U0193 -->
<s>ḡ̄̄̄ </s><!-- U029B -->
<s>ḡ̄̄̄̄ </s><!-- U0263 -->
<t>Ḡ̄̄̄̄ </t><!-- U0194 -->
<s>ḡ̄̄̄̄̄ </s><!-- U0264 -->
<s>ḡ̄̄̄̄̄̄ </s><!-- U01A3 -->
<t>Ḡ̄̄̄̄̄̄ </t><!-- U01A2 -->

<reset>h</reset> <!-- small capital h, h with hook, heng with hook, turned h,
```


turned h with fishhook, turned h with fishhook and tail -->

<s>H </s><!-- U029C -->
<s>h </s><!-- U0266 -->
<s>ĥ </s><!-- U0267 -->
<s>ȷ </s><!-- U0265 -->
<s>ɥ </s><!-- U02AE -->
<s>ɥ̣ </s><!-- U02AF -->

<reset>i</reset> <!-- dotless i, small capital i, i with stroke, latin iota -->

<s>ı </s><!-- U0131 -->
<s>İ </s><!-- U026A -->
<s>ï </s><!-- U0268 -->
<t>İ </t><!-- U0197 -->
<s>ι </s><!-- U0269 -->
<t>ι </t><!-- U0196 -->

<reset>j</reset> <!-- j with crossed tail, dotless j with stroke, dotless j with stroke and hook -->

<s>ĵ </s><!-- U029D -->
<s>ȶ </s><!-- U025F -->
<s>ƒ </s><!-- U0284 -->

<reset>k</reset> <!-- k with hook, kra, turned k -->

<s>ƀ </s><!-- U0199 -->
<t>ƀ </t><!-- U0198 -->
<s>ƶ </s><!-- U0138 -->
<s>Ʒ </s><!-- U029E -->

<reset>l</reset> <!-- small capital l, l with bar, l with middle tilde, l with belt, l with retroflex hook, l with curl, l with curl, latin lambda with stroke -->

<!-- U029F -->
<s>Ł </s><!-- U019A -->
<t>Ł </t><!-- U023D -->
<s>ł </s><!-- U026B -->
<s>ł̃ </s><!-- U026C -->
<s>ł̣ </s><!-- U026D -->
<s>ł̥ </s><!-- U0234 -->
<s>λ̣ </s><!-- U019B -->

<reset>m</reset> <!-- m with hook, turned m, turned m with long leg -->

<s>ṃ </s><!-- U0271 -->
<s>Ṃ </s><!-- U026F -->
<t>Ṃ </t><!-- U019C -->
<s>ṣ </s><!-- U0270 -->

<reset>n</reset> <!-- n preceded by apostrophe, small capital n, n with left hook, n with long right leg, n with retroflex hook, n with curl, eng -->

<s>ñ </s><!-- U0149 -->
<s>Ñ </s><!-- U0274 -->
<s>ņ </s><!-- U0272 -->
<t>Ñ </t><!-- U019D -->
<s>ņ̣ </s><!-- U019E -->

<t>Ŋ</t><!-- U0220 -->
<s>ŋ</s><!-- U0273 -->
<s>ŋ</s><!-- U0235 -->
<s>ŋ</s><!-- U014B -->
<t>Ŋ</t><!-- U014A -->

<reset>o</reset> <!-- open o, o with middle tilde, closed omega, ou -->
<s>ö</s><!-- U0254 -->
<t>Ɔ</t><!-- U0186 -->
<s>ø</s><!-- U0275 -->
<t>Θ</t><!-- U019F -->
<s>ω</s><!-- U0277 -->
<s>8</s><!-- U0223 -->
<t>8</t><!-- U0222 -->

<reset>p</reset> <!-- p with hook, latin phi -->
<s>þ</s><!-- U01A5 -->
<t>P</t><!-- U01A4 -->
<s>ϕ</s><!-- U0278 -->

<reset>q</reset> <!-- q with hook -->
<s>q̇</s><!-- U02A0 -->

<reset>r</reset> <!-- small capital r, yr, turned r, turned r with long leg,
turned r with hook, r with long leg, r with tail, r with fishhook,
reversed r with fishhook, small capital inverted r -->
<s>R</s><!-- U0280 -->
<t>Ŕ</t><!-- U01A6 -->
<s>ɹ</s><!-- U0279 -->
<s>ɹ</s><!-- U027A -->
<s>ɹ</s><!-- U027B -->
<s>ɹ</s><!-- U027C -->
<s>ɹ</s><!-- U027D -->
<s>ɹ</s><!-- U027E -->
<s>ɹ</s><!-- U027F -->
<s>ɹ</s><!-- U0281 -->

<reset>s</reset> <!-- s with hook, esh, esh loop, reversed esh,
esh with curl -->
<s>š</s><!-- U0282 -->
<s>ſ</s><!-- U0283 -->
<t>Σ</t><!-- U01A9 -->
<s>ſ</s><!-- U01AA -->
<s>ſ</s><!-- U0285 -->
<s>ſ</s><!-- U0286 -->

<reset>t</reset> <!-- t with stroke, t with palatal hook, t with hook,
t with retroflex hook, t with curl, turned t, digraph tc with curl -->
<s>ţ</s><!-- U0167 -->
<t>Ƨ</t><!-- U0166 -->
<s>ţ</s><!-- U01AB -->
<s>Ƨ</s><!-- U01AD -->

```

<t>T </t><!-- U01AC -->
<s>Ŧ </s><!-- U0288 -->
<t>Ŧ </t><!-- U01AE -->
<s>Ŧ </s><!-- U0236 -->
<s>Ŧ </s><!-- U0287 -->

<reset>u</reset> <!-- u bar, latin upsilon -->
<s>ū </s><!-- U0289 -->
<t>ū </t><!-- U0244 -->
<s>υ </s><!-- U028A -->
<t>υ </t><!-- U01B1 -->

<reset>v</reset> <!-- v with hook, turned v -->
<s>v̄ </s><!-- U028B -->
<t>ῡ </t><!-- U01B2 -->
<s>Λ </s><!-- U028C -->
<t>Λ </t><!-- U0245 -->

<reset>w</reset> <!-- turned w, wynn -->
<s>w̄ </s><!-- U028D -->
<s>ƿ </s><!-- U01BF -->
<t>ƿ </t><!-- U01F7 -->

<reset>y</reset> <!-- small capital y, y with hook, turned y, yogh -->
<s>Y </s><!-- U028F -->
<s>ȳ </s><!-- U01B4 -->
<t>Ȳ </t><!-- U01B3 -->
<s>ƿ̄ </s><!-- U028E -->
    <s>ȝ </s><!-- U021D -->
<t>ȝ </t><!-- U021C -->

<reset>z</reset> <!-- z with stroke, z with hook, z with retroflex hook,
    z with curl, ezh, ezh with caron,
    ezh reversed, ezh with tail, ezh with curl -->
<s>z̄ </s><!-- U01B6 -->
<t>Z̄ </t><!-- U01B5 -->
<s>ꝛ </s><!-- U0225 -->
<t>ꝛ̄ </t><!-- U0224 -->
<s>ꝛ̄ </s><!-- U0290 -->
<s>ꝛ̄ </s><!-- U0291 -->
<s>ȝ </s><!-- U0292 -->
<t>ȝ </t><!-- U01B7 -->
<s>ȝ̄ </s><!-- U01EF -->
<t>ȝ̄ </t><!-- U01EE -->
<s>ꝛ̄ </s><!-- U01B9 -->
<t>ꝛ̄ </t><!-- U01B8 -->
<s>ȝ̄ </s><!-- U01BA -->
<s>ȝ̄ </s><!-- U0293 -->

<!-- Digraphs -->
<reset>dz</reset> <!-- dz digraph with curl -->

```

```
<s>dz</s><!-- U02A5 -->
<reset>dʒ </reset> <!-- dezh digraph -->
<s>dʒ</s><!-- U02A4 -->
<reset>hv</reset> <!-- hv -->
<s>hv</s><!-- U0195 -->
<reset>HV</reset> <!-- hwair -->
<s>Hj</s><!-- U01F6 -->
<reset>lʒ </reset> <!-- lezh -->
<s>lʒ</s><!-- U026E -->
<reset>oe</reset> <!-- small capital oe -->
<s>œ </s><!-- U0276 -->
<reset>tʂ </reset> <!-- te digraph with curl -->
<s>tʂ</s><!-- U02A8 -->
```

<!-- tailorings against Greek obsolete when done against current UCA table -->

<!-- Cyrillic letters: full conformance with GOST requirements -->

```
<reset>ђ </reset> <!-- gje as variant of dje (Serbian) -->
```

```
<s>ђ </s><!-- U0453 -->
```

```
<t>ђ </t><!-- U0403 -->
```

```
<reset>ћ </reset> <!-- kje as variant of tshe (Serbian) -->
```

```
<s>ќ </s><!-- U045C -->
```

```
<t>ќ </t><!-- U040C -->
```

<!-- Georgian: unchanged -->

<!-- Armenian -->

```
<reset>p </reset> <!-- U0584 -->
```

```
<p>ւ </p><!-- U0587 -->
```

```
</rules>
```

```
</collation>
```

```
</collations>
```

```
</ldml>
```

Bibliography

- [1] ISO/IEC 15897:1999, *Information technology — Procedures for registration of cultural elements*
- [2] Unicode Technical Standard #35: Unicode Locale Data Markup Language (LDML). Version 1.8 (2010-04-28)
- [3] Unicode Technical Standard #10: Unicode Collation Algorithm. Version 5.2.0 (2009-10-08)
- [4] CAN/CSA Z243.230-1998, *Minimum Canadian Software Localization Conventions*

British Standards Institution (BSI)

BSI is the national body responsible for preparing British Standards and other standards-related publications, information and services.

BSI is incorporated by Royal Charter. British Standards and other standardization products are published by BSI Standards Limited.

About us

We bring together business, industry, government, consumers, innovators and others to shape their combined experience and expertise into standards-based solutions.

The knowledge embodied in our standards has been carefully assembled in a dependable format and refined through our open consultation process. Organizations of all sizes and across all sectors choose standards to help them achieve their goals.

Information on standards

We can provide you with the knowledge that your organization needs to succeed. Find out more about British Standards by visiting our website at bsigroup.com/standards or contacting our Customer Services team or Knowledge Centre.

Buying standards

You can buy and download PDF versions of BSI publications, including British and adopted European and international standards, through our website at bsigroup.com/shop, where hard copies can also be purchased.

If you need international and foreign standards from other Standards Development Organizations, hard copies can be ordered from our Customer Services team.

Subscriptions

Our range of subscription services are designed to make using standards easier for you. For further information on our subscription products go to bsigroup.com/subscriptions.

With **British Standards Online (BSOL)** you'll have instant access to over 55,000 British and adopted European and international standards from your desktop. It's available 24/7 and is refreshed daily so you'll always be up to date.

You can keep in touch with standards developments and receive substantial discounts on the purchase price of standards, both in single copy and subscription format, by becoming a **BSI Subscribing Member**.

PLUS is an updating service exclusive to BSI Subscribing Members. You will automatically receive the latest hard copy of your standards when they're revised or replaced.

To find out more about becoming a BSI Subscribing Member and the benefits of membership, please visit bsigroup.com/shop.

With a **Multi-User Network Licence (MUNL)** you are able to host standards publications on your intranet. Licences can cover as few or as many users as you wish. With updates supplied as soon as they're available, you can be sure your documentation is current. For further information, email bsmusales@bsigroup.com.

BSI Group Headquarters

389 Chiswick High Road London W4 4AL UK

Revisions

Our British Standards and other publications are updated by amendment or revision.

We continually improve the quality of our products and services to benefit your business. If you find an inaccuracy or ambiguity within a British Standard or other BSI publication please inform the Knowledge Centre.

Copyright

All the data, software and documentation set out in all British Standards and other BSI publications are the property of and copyrighted by BSI, or some person or entity that owns copyright in the information used (such as the international standardization bodies) and has formally licensed such information to BSI for commercial publication and use. Except as permitted under the Copyright, Designs and Patents Act 1988 no extract may be reproduced, stored in a retrieval system or transmitted in any form or by any means – electronic, photocopying, recording or otherwise – without prior written permission from BSI. Details and advice can be obtained from the Copyright & Licensing Department.

Useful Contacts:

Customer Services

Tel: +44 845 086 9001

Email (orders): orders@bsigroup.com

Email (enquiries): cservices@bsigroup.com

Subscriptions

Tel: +44 845 086 9001

Email: subscriptions@bsigroup.com

Knowledge Centre

Tel: +44 20 8996 7004

Email: knowledgecentre@bsigroup.com

Copyright & Licensing

Tel: +44 20 8996 7070

Email: copyright@bsigroup.com



...making excellence a habit.™