

---

---

**Alphabetical ordering of multilingual  
terminological and lexicographical  
data represented in the Latin alphabet**

*Mise en ordre alphabétique des données lexicographiques et  
terminologiques multilingues représentées dans l'alphabet latin*

STANDARDSISO.COM : Click to view the full PDF of ISO 12199:2022



STANDARDSISO.COM : Click to view the full PDF of ISO 12199:2022



**COPYRIGHT PROTECTED DOCUMENT**

© ISO 2022

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office  
CP 401 • Ch. de Blandonnet 8  
CH-1214 Vernier, Geneva  
Phone: +41 22 749 01 11  
Email: [copyright@iso.org](mailto:copyright@iso.org)  
Website: [www.iso.org](http://www.iso.org)

Published in Switzerland

# Contents

Page

|                                                                                                      |           |
|------------------------------------------------------------------------------------------------------|-----------|
| <b>Foreword</b> .....                                                                                | <b>iv</b> |
| <b>Introduction</b> .....                                                                            | <b>v</b>  |
| <b>1 Scope</b> .....                                                                                 | <b>1</b>  |
| <b>2 Normative references</b> .....                                                                  | <b>1</b>  |
| <b>3 Terms and definitions</b> .....                                                                 | <b>1</b>  |
| <b>4 Preparatory procedures</b> .....                                                                | <b>2</b>  |
| <b>5 First ordering level</b> .....                                                                  | <b>3</b>  |
| 5.1 First-ordering-level values.....                                                                 | 3         |
| 5.2 First-ordering-level sequence.....                                                               | 3         |
| 5.3 Equivalence between special Latin letters and basic letters.....                                 | 4         |
| <b>6 Second ordering level</b> .....                                                                 | <b>4</b>  |
| 6.1 Second-ordering-level values.....                                                                | 4         |
| 6.2 Special Latin letters and letters with diacritical marks.....                                    | 5         |
| <b>7 Third ordering level</b> .....                                                                  | <b>6</b>  |
| 7.1 Third-ordering-level values.....                                                                 | 6         |
| 7.2 Ordering according to capitalization.....                                                        | 6         |
| <b>8 Fourth ordering level</b> .....                                                                 | <b>6</b>  |
| 8.1 Fourth-ordering-level values.....                                                                | 6         |
| 8.2 Ordering according to special characters.....                                                    | 6         |
| <b>Annex A (normative) Word-by-word ordering</b> .....                                               | <b>7</b>  |
| <b>Annex B (informative) Special rules for lexicographical and terminological ordering</b> .....     | <b>9</b>  |
| <b>Annex C (informative) Ordering rules for chemical names</b> .....                                 | <b>10</b> |
| <b>Annex D (informative) Character repertoire of the Latin alphabet</b> .....                        | <b>12</b> |
| <b>Annex E (informative) Languages using the Latin alphabet</b> .....                                | <b>20</b> |
| <b>Annex F (informative) Alphabetical sequences and character repertoires</b> .....                  | <b>27</b> |
| <b>Annex G (informative) Formal description of the rules of the main body of this document</b> ..... | <b>40</b> |
| <b>Bibliography</b> .....                                                                            | <b>50</b> |

## Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see [www.iso.org/directives](http://www.iso.org/directives)).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see [www.iso.org/patents](http://www.iso.org/patents)).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see [www.iso.org/iso/foreword.html](http://www.iso.org/iso/foreword.html).

This document was prepared by Technical Committee ISO/TC 37, *Language and terminology*, Subcommittee SC 2, *Terminology workflow and language coding*.

This second edition cancels and replaces the first edition (ISO 12199:2000), of which it constitutes a minor revision. The changes are as follows:

- the relationship of this document with other International Standards has been updated and transferred from the Foreword to the Introduction;
- in [Clause 2](#) and in the Bibliography, the references have been updated;
- ISO/IEC 14651 is cited informatively and therefore has been moved from [Clause 2](#) to the Bibliography;
- in [Annexes D, E and F](#), the Serbian language has been added among the languages using the Latin alphabet, together with a character set and alphabetical ordering information relating to the Serbian language;
- in [Annex E](#), the references to Serbo-Croatian have been deleted;
- in [Annexes E and F](#), the entries related to Moldovan have been corrected in line with ISO 639-1 and ISO 639-2;
- [Annex G](#) is cited informatively and therefore has been changed to “(informative)”.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at [www.iso.org/members.html](http://www.iso.org/members.html).

## Introduction

In the development of international terminologies, both in printed form and in databases, it is essential to have uniform and internationally recognized rules for the alphabetical ordering of terminological and lexicographical data, to make these terminologies more easily accessible for the users. In addition, it will facilitate the interchange of terminological and lexicographical data.

This document complements other International Standards, such as ISO 10241-1.

STANDARDSISO.COM : Click to view the full PDF of ISO 12199:2022

STANDARDSISO.COM : Click to view the full PDF of ISO 12199:2022

# Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet

## 1 Scope

This document specifies the sequence of characters to be used in the alphabetical ordering of multilingual terminological and lexicographical data (terms, term elements, or words) represented in the Latin alphabet. Character sets of languages represented in the Latin alphabet are taken into account insofar as terminological or lexicographical data have been recorded. Character sets used in internationally standardized transliteration into Latin script are also taken into account.

The sequence of alphabetical characters given is intended for multilingual purposes only and is not intended to affect the alphabetical order of any specific language.

The main part of this document specifies letter-by-letter ordering of character strings. [Annex A](#) treats word-by-word ordering, which is a widely used alternative to this system.

[Annex B](#) gives two additional rules that can be useful for lexicographical and terminological ordering.

[Annex C](#) gives ordering rules for chemical names.

[Annex D](#) lists the character repertoire of the Latin alphabet.

[Annex E](#) lists languages using the Latin alphabet.

[Annex F](#) gives alphabetical sequences derived from the sequence specified in this document for a number of languages that use the Latin alphabet.

[Annex G](#) gives a formal description of the rules laid down in the main part of this document conforming with ISO/IEC 14651.

## 2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements 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 1087, *Terminology work and terminology science — Vocabulary*

ISO/IEC 10646-1<sup>1)</sup>, *Information technology — Universal Multiple-Octet Coded Character Set (UCS) — Part 1: Architecture and Basic Multilingual Plane*

## 3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 1087 and the following apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

- ISO Online browsing platform: available at <https://www.iso.org/obp>
- IEC Electropedia: available at <https://www.electropedia.org/>

1) In this minor revision of ISO 12199:2000, reference continues to be made to ISO/IEC 10646-1:1993. ISO/IEC 10646-1 and ISO/IEC 10646-2 have since been merged into ISO/IEC 10646:2020.

**3.1**  
**character**

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

**3.2**  
**letter**

*character* (3.1) used for writing natural language, often representing a sound in the language

**3.3**  
**digit**

*character* (3.1) used to represent the numeric value, or part thereof, of a number

**3.4**  
**special character**

*character* (3.1) that is not a *letter* (3.2) nor a *digit* (3.3)

EXAMPLE The space character is a special character.

**3.5**  
**ligature**

*character* (3.1) resulting from the joining of two or more *letters* (3.2)

Note 1 to entry: The resulting character is, in some cases, considered a separate letter.

**3.6**  
**polygraph**

two or more consecutive *letters* (3.2) that are regarded as one letter for some purpose

Note 1 to entry: A polygraph consisting of two or three letters may be referred to as a digraph or a trigraph, respectively.

**3.7**  
**diacritical mark**

*character* (3.1) that is not a *letter* (3.2) and is placed over, under, or through a letter or a combination of letters

**3.8**  
**ordering**

act of bringing strings of *characters* (3.1) into a well-defined sequence according to a string comparison specification

## 4 Preparatory procedures

In the process of alphabetical ordering, character strings are compared according to a set of rules. This document specifies the set of rules to be used for the ordering, but does not address the means of selection of relevant character strings, nor any modification of the strings that can be needed for a given purpose. Consequently, certain preparatory procedures can be needed before applying the ordering rules. Depending on the needs in each individual case, it is possible that:

- the relevant character strings have to be selected, e.g. relevant terms have to be extracted from a corpus;
- the character strings have to be modified, e.g. sentence-initial uppercase letters have to be changed to lowercase letters, plural form of words have to be changed to singular form;
- leading zeroes or spaces can be added, e.g. in lists containing numerals.

Polygraphs are treated as sequences of separate letters.



An application may arrange information into several ordering fields, and determine ranking order with several separate and independent comparisons. This document only defines a single comparison for one such field, where the field is a character-string field.

Only the characters that appear in the string and their arrangement are taken into account. Apart from the ordering rules and passes, no other knowledge about the words in the character string is used. For example, dictionary information or rules about language syntax, phonetics and semantics are not used.

## 5 First ordering level

### 5.1 First-ordering-level values

When comparing strings to be ordered, the first-ordering-level values of the strings shall be considered first. The subsequent ordering-level values need to be considered only if two or more strings have identical first-ordering-level values.

For multilingual ordering, the following rules shall be applied ([Annex A](#) shall be applied for word-by-word ordering).

### 5.2 First-ordering-level sequence

Digits and letters have the following ordering values:

#### a) Digits:

0 1 2 3 4 5 6 7 8 9

NOTE 1 Sequences of digits are ordered from left to right as written, thus generating the following order, for example: 1 10 100 11 110 111 12 19 190 2 21 3.

NOTE 2 Leading zeroes can be inserted as a preparatory procedure, e.g. to generate the following order: 0001 0002 0003 0010 0011 0012 0019 0021 0100 0110 0111 0190.

#### b) Basic letters of the Latin alphabet:

a A b B c C d D e E f F g G h H i I j J k K l L m M n N  
o O p P q Q r R s S t T u U v V w W x X y Y z Z þ Þ

NOTE 3 This order has been established for use in multilingual environments so as to conflict with as few individual languages as possible. See [Annex F](#) for examples of deviations from this sequence in some languages.

Uppercase and lowercase letters shall be treated as equivalent (see [Clause 7](#)). Letters of the Latin alphabet with diacritical marks shall be treated as equivalent to the corresponding basic Latin letters (see [Clause 6](#)). Special letters of the Latin alphabet shall be treated as equivalent to basic Latin letters according to [Table 1](#) in [5.3](#) (see [Clause 6](#)).

The Turkish language distinguishes i/I from i/İ, while other languages have the pair i/I only. To order multilingual data including Turkish text, the i/I pair shall be expanded as follows:

- 1: i/I U0131/U0049 LATIN LETTER DOTLESS I (Turkish)
- 2: i/I U0131/U0049 LATIN LETTER I (non-Turkish)
- 3: i/İ U0069/U0130 LATIN LETTER I WITH DOT ABOVE (Turkish)

It should also be noted that, for example, í (U00ED LATIN SMALL LETTER I WITH ACUTE) in normal print is represented as LATIN SMALL LETTER DOTLESS I WITH ACUTE. For the purpose of ordering, however, it shall be treated as equivalent to i (U0069 LATIN SMALL LETTER I) on the first ordering level.

NOTE 4 Throughout this document, characters are referenced as UXXXX, where X is any hexadecimal digit and refers to the position of the character in ISO/IEC 10646-1. Character names are given as in ISO/IEC 10646-1. Most names of Latin letters start with “LATIN SMALL LETTER ...” and “LATIN CAPITAL LETTER ...”. When referring to both lowercase and uppercase letter, the name “LATIN LETTER ...” is used. When there is no danger of misinterpretation, the words “LATIN LETTER” are sometimes omitted.

### c) Letters of other alphabets:

Letters of other alphabets follow in the sequences established for each alphabet. The order of non-Latin alphabets shall be: the Greek alphabet, the Cyrillic alphabet, other alphabets.

NOTE 5 It is outside the scope of this document to establish the sequences for alphabets other than the Latin alphabet. The Greek alphabet has the following sequence of letters:

α Α β Β γ Γ δ Δ ε Ε ζ Ζ η Η θ Θ ι Ι κ Κ λ Λ μ Μ ν Ν ξ Ξ  
ο Ο π Π ρ Ρ σ Σ τ Τ υ Υ φ Φ χ Χ ψ Ψ ω Ω

All other characters, e.g. punctuation marks, shall be ignored. See [Clause 8](#).

## 5.3 Equivalence between special Latin letters and basic letters

Special Latin letters shall be treated as equivalent to basic letters of the Latin alphabet according to [Table 1](#). Uppercase and lowercase letters shall be treated as equivalent.

**Table 1 — Equivalence between special Latin letters and basic letters**

| Position | Character name in ISO/IEC 10646-1 | Character position for lowercase/uppercase in ISO/IEC 10646-1 |              | Equivalent to |
|----------|-----------------------------------|---------------------------------------------------------------|--------------|---------------|
| 01       | LATIN LETTER AE                   | U00E6                                                         | U00C6        | ae            |
| 02       | LATIN LETTER B WITH HOOK          | U0253                                                         | U0181        | b             |
| 03       | LATIN LETTER C WITH HOOK          | U0188                                                         | U0187        | c             |
| 04       | LATIN LETTER D WITH STROKE        | U0111                                                         | U0110        | d             |
| 05       | LATIN LETTER D WITH HOOK          | U0257                                                         | U018A        | d             |
| 06       | LATIN LETTER ETH                  | U00F0                                                         | U00D0        | d             |
| 07       | LATIN LETTER G WITH HOOK          | U0260                                                         | U0193        | g             |
| 08       | LATIN LETTER H WITH STROKE        | U0127                                                         | U0126        | h             |
| 09       | LATIN LETTER K WITH HOOK          | U0199                                                         | U0198        | k             |
| 10       | LATIN SMALL LETTER KRA            | U0138                                                         | <sup>a</sup> | k             |
| 11       | LATIN LETTER L WITH STROKE        | U0142                                                         | U0141        | l             |
| 12       | LATIN LETTER ENG                  | U014B                                                         | U014A        | n             |
| 13       | LATIN LETTER O WITH STROKE        | U00F8                                                         | U00D8        | o             |
| 14       | LATIN LIGATURE OE                 | U0153                                                         | U0152        | oe            |
| 15       | LATIN SMALL LETTER SHARP S        | U00DF                                                         | <sup>a</sup> | ss            |
| 16       | LATIN LETTER T WITH STROKE        | U0167                                                         | U0166        | t             |

<sup>a</sup> No corresponding uppercase letter.

## 6 Second ordering level

### 6.1 Second-ordering-level values

If the comparison of two strings results in identical first-ordering-level values, second-ordering-level values shall be applied according to [6.2](#).

The rule shall be applied from left to right.

## 6.2 Special Latin letters and letters with diacritical marks

Special Latin letters, that have been treated as equivalent to basic Latin letters according to [Table 1](#), shall be ordered according to the order in [Table 1](#).

Diacritical marks shall be ordered according to [Table 2](#).

NOTE This order has been established for multilingual environments so as to be in conflict with as few individual languages as possible. See [Annex F](#) for examples of deviations from this sequence in some languages.

**Table 2 — Ordering of diacritical marks**

| Position                                                                         | Name                      | Position for combining diacritical mark in ISO/IEC 10646-1 |
|----------------------------------------------------------------------------------|---------------------------|------------------------------------------------------------|
| 0000                                                                             | none                      | —                                                          |
| 0100                                                                             | ACUTE ACCENT              | U0301                                                      |
| 0200                                                                             | GRAVE ACCENT              | U0300                                                      |
| 0300                                                                             | BREVE                     | U0306                                                      |
| 0301                                                                             | BREVE AND ACUTE           | —                                                          |
| 0302                                                                             | BREVE AND GRAVE           | —                                                          |
| 0310                                                                             | BREVE AND HOOK ABOVE      | —                                                          |
| 0311                                                                             | BREVE AND TILDE           | —                                                          |
| 0313                                                                             | BREVE AND DOT BELOW       | —                                                          |
| 0315                                                                             | BREVE AND COMMA BELOW     | —                                                          |
| 0400                                                                             | CIRCUMFLEX ACCENT         | U0302                                                      |
| 0401                                                                             | CIRCUMFLEX AND ACUTE      | —                                                          |
| 0402                                                                             | CIRCUMFLEX AND GRAVE      | —                                                          |
| 0410                                                                             | CIRCUMFLEX AND HOOK ABOVE | —                                                          |
| 0411                                                                             | CIRCUMFLEX AND TILDE      | —                                                          |
| 0413                                                                             | CIRCUMFLEX AND DOT BELOW  | —                                                          |
| 0500                                                                             | CIRCUMFLEX ACCENT BELOW   | U032D                                                      |
| 0600                                                                             | CARON                     | U030C                                                      |
| 0614                                                                             | CARON AND CEDILLA         | —                                                          |
| 0700                                                                             | RING ABOVE                | U030A                                                      |
| 0701                                                                             | RING ABOVE AND ACUTE      | —                                                          |
| 0800                                                                             | DIAERESIS                 | U0308                                                      |
| 0813                                                                             | DIAERESIS AND DOT BELOW   | —                                                          |
| 0817                                                                             | DIAERESIS AND MACRON      | —                                                          |
| 0900                                                                             | DOUBLE ACUTE ACCENT       | U030B                                                      |
| 1000                                                                             | HOOK ABOVE                | U0309                                                      |
| 1100                                                                             | TILDE                     | U0303                                                      |
| 1200                                                                             | DOT ABOVE                 | U0307                                                      |
| 1300                                                                             | DOT BELOW                 | U0323                                                      |
| 1400                                                                             | CEDILLA                   | U0327                                                      |
| 1500                                                                             | COMMA ABOVE/BELOW         | U0313 and U0326 <sup>a</sup>                               |
| 1600                                                                             | OGONEK                    | U0328                                                      |
| <sup>a</sup> The position of combining comma above and below the base character. |                           |                                                            |

**Table 2** (continued)

| Position                                                                         | Name                   | Position for combining diacritical mark in ISO/IEC 10646-1 |
|----------------------------------------------------------------------------------|------------------------|------------------------------------------------------------|
| 1700                                                                             | MACRON                 | U0304                                                      |
| 1713                                                                             | MACRON AND DOT BELOW   | —                                                          |
| 1800                                                                             | MACRON BELOW           | U0331                                                      |
| 1900                                                                             | PRECEDED BY APOSTROPHE | —                                                          |
| 2000                                                                             | FOLLOWED BY APOSTROPHE | —                                                          |
| 2100                                                                             | HORN                   | U031B                                                      |
| 2101                                                                             | HORN AND ACUTE         | —                                                          |
| 2102                                                                             | HORN AND GRAVE         | —                                                          |
| 2110                                                                             | HORN AND HOOK ABOVE    | —                                                          |
| 2111                                                                             | HORN AND TILDE         | —                                                          |
| 2113                                                                             | HORN AND DOT BELOW     | —                                                          |
| <sup>a</sup> The position of combining comma above and below the base character. |                        |                                                            |

## 7 Third ordering level

### 7.1 Third-ordering-level values

If the comparison of two strings results in identical first- and second-ordering-level values, third-ordering-level values shall be applied according to [7.2](#).

The rule shall be applied from left to right.

### 7.2 Ordering according to capitalization

A lowercase letter shall be ordered before the corresponding uppercase letter. [See [5.2](#), item b), first paragraph after NOTE 3.]

NOTE The terms “lowercase letter” and “uppercase letter” are used for members of the sets “a b c ...” and “A B C ...”, respectively. In character names, the naming conventions of ISO/IEC 10646-1 are used. ISO/IEC 10646-1 uses “LATIN SMALL LETTER” and “LATIN CAPITAL LETTER”, respectively.

## 8 Fourth ordering level

### 8.1 Fourth-ordering-level values

If the comparison of two strings results in identical first-, second- and third-ordering-level values, fourth-ordering-level values shall be applied according to [8.2](#).

The rule shall be applied from left to right.

### 8.2 Ordering according to special characters

Special characters are ordered according to the sequence of the default template of ISO/IEC 14651. For most special characters, this is the order in which they are listed in ISO/IEC 10646-1.

NOTE In word-by-word ordering (see [Annex A](#)), the space character and possibly other special characters can have special functions as key separators.

## Annex A (normative)

### Word-by-word ordering

#### A.1 Principles of word-by-word ordering

As noted in the Scope, this document specifies the letter-by-letter ordering of character strings. Word-by-word ordering is a widely used alternative to this system. [Table A.1](#) illustrates the difference between letter-by-letter ordering and word-by-word ordering.

**Table A.1 — Letter-by-letter and word-by-word ordering**

| Letter-by-letter ordering | Word-by-word ordering |
|---------------------------|-----------------------|
| ad                        | ad                    |
| adhesive                  | ad hoc                |
| ad hoc                    | ad infinitum          |
| adieu                     | adhesive              |
| ad infinitum              | adieu                 |
| adipose                   | adipose               |

#### A.2 Multiple-key ordering

Single-key ordering is described in the main body of this document. In multiple-key ordering, all the ordering rules are applied to one key before they are applied to the next, until all the keys have been considered or a unique sequence has been established.

**NOTE** One typical example of multiple-key ordering is a list of delegates to a meeting, where the first key can be the country names, the second key can be the delegates' last names, and the third key can be the delegates' first names. In this example, if a country has one delegate only, the second key (last names) will not be considered.

#### A.3 Word-by-word ordering as multiple-key ordering

In word-by-word ordering, space characters, and possibly also by definition other characters, are key separators. The key-separator characters function as key separators only, and they have no position in the ordering sequence.

When the character string has been divided into a sequence of keys, the ordering rules of the main body of this document are invoked for one key at a time.

**NOTE 1** In addition to the space characters, some or all punctuation marks can be defined as key separators. It can also be useful to define some space characters as key separators, while other space characters remain special characters within a key. The choices depend on the language(s) and type of strings to be ordered.

**NOTE 2** If space characters and hyphens are defined as key separators, the title of this clause would be split into the following keys: <A.3> <Word> <by> <word> <ordering> <as> <multiple> <key> <ordering>, where each key is contained within < and >, and the spaces are added for increased readability.

#### A.4 Simple word-by-word ordering

If the text to be ordered using word-by-word ordering contains very few special Latin letters and diacritical marks, the following extension to the rules in the main body of this document will produce the same or nearly the same output as the rules described in [Clause A.3](#).

On the first ordering level (see [5.2](#)), the **space character** is added as the first item. Items 1, 2, and 3 in [5.2](#) then become items 2, 3, and 4. The space character is not treated as a special character on the fourth ordering level (see [Clause 8](#)).

NOTE Depending on the language(s) and type of strings to be ordered, it can be useful to treat even other special characters (e.g. hyphens) in the same way as the space character.

STANDARDSISO.COM : Click to view the full PDF of ISO 12199:2022

## Annex B (informative)

### Special rules for lexicographical and terminological ordering

#### B.1 Background

For lexicographical and terminological applications, it can sometimes be desirable to add additional ordering criteria to the rules that are described in the main body of this document.

The features that are described in this annex cannot easily be described in the formalism given in ISO/IEC 14651.

#### B.2 Position relative to baseline

It can be desirable to distinguish, for example,  $m_2$ ,  $m^2$ ,  $m_2$  for ordering purposes. If this is deemed necessary, it is recommended that this be done on the third ordering level (see [Clause 7](#)) combined with capitalization.

The ordering value of any given character based on its position relative to the baseline may be determined according to [Table B.1](#).

**Table B.1 — Position relative to baseline**

|   |                                                       |
|---|-------------------------------------------------------|
| 1 | character(s) on baseline                              |
| 2 | character(s) above baseline, superscript character(s) |
| 3 | character(s) below baseline, subscript character(s)   |

#### B.3 Ordering according to styles

If ordering by the first through fourth ordering level does not produce a unique sequence, typographical styles may be taken into consideration as a fifth ordering level.

Styles may be ordered according to [Table B.2](#).

**Table B.2 — Order of styles**

| Position | Style name      | Example                    |
|----------|-----------------|----------------------------|
| 1        | roman           | abcdefghijkl               |
| 2        | boldface        | <b>abcdefghijkl</b>        |
| 3        | italic          | <i>abcdefghijkl</i>        |
| 4        | boldface-italic | <b><i>abcdefghijkl</i></b> |
| 5        | others          |                            |

## Annex C (informative)

### Ordering rules for chemical names

#### C.1 Background

There are no universally accepted ordering rules for chemical names. The ordering rules of the main body of this document may be used, if so desired, with the extension of the word-by-word ordering rules described in [Annex A](#).

However, some indexes and databases, in particular at the Chemical Abstracts Services (CAS), use a specially designed multiple-key ordering system. The main features of this system are outlined in this annex.<sup>2)</sup>

#### C.2 Division into three keys

##### C.2.1 Parent name

The first key consists of the parent name, which normally is all roman letters and space characters, whether or not interrupted by italic letters, Greek letters, digits or special characters (e.g. punctuation).

##### C.2.2 Initial locants

The second key consists of initial locants, being all characters before the first roman letter.

##### C.2.3 Other locants

The third key consists of all non-initial locants, being all remaining characters.

NOTE The name “2-Butanone-1,1,1-*d*<sub>3</sub>, 3,3-dimethyl” is divided into three keys as follows: <Butanone dimethyl> <2-> <-1,1,1-*d*<sub>3</sub>, 3,3->.

#### C.3 Ordering rules within each key

The first key is ordered according to the rules of the main body of this document.

In the second and third keys, the following order is used:

- letters of the Latin alphabet (which is in italic), in the order specified in [5.2](#), item b);
- letter of the Greek alphabet, in the order given in [5.2](#), item c);
- numerals, in the order of the numeric value.

#### C.4 Output

[Table C.1](#) shows ordered output from the rules that are described in this annex compared with output from the rules of the main body of this document.

2) For further details, consult Chemical Abstracts Services (CAS), P.O. Box 3012, Columbus, Ohio 43210, USA.



Table C.1 — Sample output

| Ordered according to Annex C                            | Ordered according to general rules                      |
|---------------------------------------------------------|---------------------------------------------------------|
| Bromine fluoride (BrF <sub>3</sub> )                    | 1-Butanone                                              |
| Bromine fluoride (BrF <sub>5</sub> )                    | 1-Butanone, 1-phenyl-                                   |
| 2-Butanol                                               | 2-Butanol                                               |
| 2-Butanol, ( <i>R</i> )-                                | 2-Butanol, 1-chloro-                                    |
| 2-Butanol, ( <i>S</i> )-                                | 2-Butanol, 4-(trimethylstannyl)-                        |
| 2-Butanol, sodium salt, ( <i>S</i> )-                   | 2-Butanol, ( <i>R</i> )-                                |
| 2-Butanol, 1-chloro-                                    | 2-Butanol, ( <i>S</i> )-                                |
| 2-Butanol, 4-(trimethylstannyl)-                        | 2-Butanol, sodium salt, ( <i>S</i> )-                   |
| 1-Butanone                                              | 2-Butanone                                              |
| 1-Butanone, 1-phenyl-                                   | 2-Butanone, 1-(dimethylamino)-3,3-dimethyl-             |
| 2-Butanone                                              | 2-Butanone-1,1,1- <i>d</i> <sub>3</sub>                 |
| 2-Butanone, <i>O</i> -methyloxime                       | 2-Butanone-1,1,1- <i>d</i> <sub>3</sub> , 3,3-dimethyl- |
| 2-Butanone, oxime                                       | 2-Butanone, 3-(4-acetylphenyl)-                         |
| 2-Butanone, polymer with formaldehyde                   | 2-Butanone, 3-ethoxy-1,1-dihydroxy-                     |
| 2-Butanone, 3-(4-acetylphenyl)-                         | 2-Butanone, <i>O</i> -methyloxime                       |
| 2-Butanone, 1-(dimethylamino)-3,3-dimethyl-             | 2-Butanone, oxime                                       |
| 2-Butanone, 3-ethoxy-1,1-dihydroxy-                     | 2-Butanone, polymer with formaldehyde                   |
| 2-Butanone-1,1,1- <i>d</i> <sub>3</sub>                 | Bromine fluoride (BrF <sub>3</sub> )                    |
| 2-Butanone-1,1,1- <i>d</i> <sub>3</sub> , 3,3-dimethyl- | Bromine fluoride (BrF <sub>5</sub> )                    |
| Butanoyl chloride                                       | Butanoyl chloride                                       |

## Annex D (informative)

### Character repertoire of the Latin alphabet

[Table D.1](#) lists the character repertoire of the Latin alphabet. The languages listed in [Annex E](#) have been taken into account if reliable information is available. Characters that are exclusive to the International Phonetic Alphabet have not been included.

NOTE The names used in ISO/IEC 10646-1 are used in the “Name” column of [Table D.1](#). The full names of the letters are “LATIN SMALL LETTER ...” and “LATIN CAPITAL LETTER ...” for the lowercase and uppercase letters, respectively. In the “Type” column of [Table D.1](#), b = basic Latin letter; d = Latin letter with diacritical mark; s = special Latin letter. In the column “languages used”, + indicates that the letter is used in most or all languages that use the Latin alphabet. For the language symbols used in [Table D.1](#), see [Annexes E](#) and [F](#). The language symbols in square brackets refer to transliteration systems; see [Table F.2](#).

**Table D.1 — Character repertoire**

| Name                              | Type | Position for<br>lowercase/uppercase<br>in ISO/IEC 10646-1 |       | Languages used                                                                     |
|-----------------------------------|------|-----------------------------------------------------------|-------|------------------------------------------------------------------------------------|
| LATIN LETTER A                    | b    | U0061                                                     | U0041 | +                                                                                  |
| WITH ACUTE                        | d    | U00E1                                                     | U00C1 | af ca cs cy da es fo fur ga gd gl<br>hu is kl nl pt qal sk smi ss vi<br>[Cyr] [ar] |
| WITH GRAVE                        | d    | U00E0                                                     | U00C0 | ca cy de fr fur fy gd it nl no pt<br>qal rm vi [Cyr]                               |
| WITH BREVE                        | d    | U0103                                                     | U0102 | mo ro vi [Cyr]                                                                     |
| WITH BREVE AND ACUTE              | d    | U1EAF                                                     | U1EAE | vi                                                                                 |
| WITH BREVE AND GRAVE              | d    | U1EB1                                                     | U1EB0 | vi                                                                                 |
| WITH BREVE AND<br>HOOK ABOVE      | d    | U1EB3                                                     | U1EB2 | vi                                                                                 |
| WITH BREVE AND TILDE              | d    | U1EB5                                                     | U1EB4 | vi                                                                                 |
| WITH BREVE AND<br>DOT BELOW       | d    | U1EB7                                                     | U1EB6 | vi                                                                                 |
| WITH CIRCUMFLEX                   | d    | U00E2                                                     | U00C2 | br cy de fr fur fy kl mo pt qal rm<br>ro smi vi [Cyr] [ar]                         |
| WITH CIRCUMFLEX AND<br>ACUTE      | d    | U1EA5                                                     | U1EA4 | vi                                                                                 |
| WITH CIRCUMFLEX AND<br>GRAVE      | d    | U1EA7                                                     | U1EA6 | vi                                                                                 |
| WITH CIRCUMFLEX AND<br>HOOK ABOVE | d    | U1EA9                                                     | U1EA8 | vi                                                                                 |
| WITH CIRCUMFLEX AND<br>TILDE      | d    | U1EAB                                                     | U1EAA | vi                                                                                 |
| WITH CIRCUMFLEX AND<br>DOT BELOW  | d    | U1EAD                                                     | U1EAC | vi                                                                                 |
| WITH CARON                        | d    | U01CE                                                     | U01CD | [Cyr]                                                                              |
| WITH RING ABOVE                   | d    | U00E5                                                     | U00C5 | da kl no smi sv [Cyr]                                                              |
| WITH RING ABOVE AND<br>ACUTE      | d    | U01FB                                                     | U01FA |                                                                                    |

Table D.1 (continued)

| Name                         | Type | Position for lowercase/uppercase in ISO/IEC 10646-1 |       | Languages used                          |
|------------------------------|------|-----------------------------------------------------|-------|-----------------------------------------|
| WITH DIAERESIS               | d    | U00E4                                               | U00C4 | cy de et fi fy lb nl sk smi sv tr [Cyr] |
| WITH DIAERESIS AND DOT BELOW | d    | —                                                   | —     | [Cyr]                                   |
| WITH DIAERESIS AND MACRON    | d    | U01DF                                               | U01DE |                                         |
| WITH DOUBLE ACUTE            | d    | —                                                   | —     | [Cyr]                                   |
| WITH HOOK ABOVE              | d    | U1EA3                                               | U1EA2 | vi                                      |
| WITH TILDE                   | d    | U00E3                                               | U00C3 | kl pt vi                                |
| WITH DOT BELOW               | d    | U1EA1                                               | U1EA0 | vi                                      |
| WITH OGONEK                  | d    | U0105                                               | U0104 | lt pl                                   |
| WITH MACRON                  | d    | U0101                                               | U0100 | lv [Cyr] [ar]                           |
| LATIN LETTER AE              | s    | U00E6                                               | U00C6 | da fo fr is kl no smi [Cyr]             |
| WITH ACUTE                   | s d  | U01FD                                               | U01FC |                                         |
| WITH MACRON                  | s d  | U01E3                                               | U01E2 |                                         |
| LATIN LETTER B               | b    | U0062                                               | U0042 | +                                       |
| WITH DOT ABOVE               | d    | U1E03                                               | U1E02 | [he]                                    |
| WITH DOT BELOW               | d    | U1E05                                               | U1E04 |                                         |
| WITH HOOK                    | s    | U0253                                               | U0181 | ha                                      |
| LATIN LETTER C               | b    | U0063                                               | U0043 | +                                       |
| WITH ACUTE                   | d    | U0107                                               | U0106 | hr pl sr [Cyr]                          |
| WITH GRAVE                   | d    | —                                                   | —     | [Cyr]                                   |
| WITH BREVE                   | d    | —                                                   | —     | [Cyr]                                   |
| WITH BREVE AND COMMA BELOW   | d    | —                                                   | —     | [Cyr]                                   |
| WITH CIRCUMFLEX              | d    | U0109                                               | U0108 | eo [Cyr]                                |
| WITH CARON                   | d    | U010D                                               | U010C | cs hr lt lv sk sl smi sr [Cyr]          |
| WITH DIAERESIS               | d    | —                                                   | —     | [Cyr]                                   |
| WITH DOT ABOVE               | d    | U010B                                               | U010A | mt                                      |
| WITH DOT BELOW               | d    | —                                                   | —     | [Cyr]                                   |
| WITH CEDILLA                 | d    | U00E7                                               | U00C7 | ca fr oc pt sq tr [Cyr]                 |
| WITH MACRON                  | d    | —                                                   | —     | [Cyr]                                   |
| WITH HOOK                    | s    | U0188                                               | U0187 |                                         |
| LATIN LETTER D               | b    | U0064                                               | U0044 | +                                       |
| WITH CIRCUMFLEX              | d    | —                                                   | —     | [Cyr]                                   |
| WITH CIRCUMFLEX BELOW        | d    | U1E13                                               | U1E12 | hz ve                                   |
| WITH CARON                   | d    | U010F                                               |       | (cs) (sk)                               |
| WITH CARON                   | d    |                                                     | U010E | cs sk                                   |
| WITH DOT ABOVE               | d    | U1E0B                                               | U1E0A |                                         |
| WITH DOT BELOW               | d    | U1E0D                                               | U1E0C | [ar]                                    |
| WITH LINE BELOW              | d    | U1E0F                                               | U1E0E | [ar]                                    |
| FOLLOWED BY APOSTROPHE       | d    | —                                                   |       | cs sk                                   |

Table D.1 (continued)

| Name                                | Type | Position for lowercase/uppercase in ISO/IEC 10646-1 |       | Languages used                                                                 |
|-------------------------------------|------|-----------------------------------------------------|-------|--------------------------------------------------------------------------------|
| WITH STROKE <sup>a</sup>            | s    | U0111                                               | U0110 | hr smi sr vi [Cyr]                                                             |
| WITH HOOK                           | s    | U0257                                               | U018A | ha                                                                             |
| LATIN LETTER ETH <sup>a</sup>       | s    | U00F0                                               | U00D0 | fo is                                                                          |
| LATIN LETTER E                      | b    | U0065                                               | U0045 | +                                                                              |
| WITH ACUTE                          | d    | U00E9                                               | U00C9 | af ca cs cy da de es fr fy ga gd gl hu is it kl lb nl no pt qal sk sl ss sv vi |
| WITH GRAVE                          | d    | U00E8                                               | U00C8 | af ca cy de fr fur gd it nl no pt qal rm vi [Cyr]                              |
| WITH BREVE                          | d    | U0115                                               | U0114 | [Cyr]                                                                          |
| WITH CIRCUMFLEX                     | d    | U00EA                                               | U00CA | af br cy de fr fy nl no nso pt qal rm sl tn vi [Cyr]                           |
| WITH CIRCUMFLEX AND ACUTE           | d    | U1EBF                                               | U1EBE | vi                                                                             |
| WITH CIRCUMFLEX AND GRAVE           | d    | U1EC1                                               | U1EC0 | vi                                                                             |
| WITH CIRCUMFLEX AND HOOK ABOVE      | d    | U1EC3                                               | U1EC2 | vi                                                                             |
| WITH CIRCUMFLEX AND TILDE           | d    | U1EC5                                               | U1EC4 | vi                                                                             |
| WITH CIRCUMFLEX AND DOT BELOW       | d    | U1EC7                                               | U1EC6 | vi                                                                             |
| WITH CIRCUMFLEX BELOW               | d    | U1E19                                               | U1E18 |                                                                                |
| WITH CARON                          | d    | U011B                                               | U011A | cs [Cyr]                                                                       |
| WITH DIAERESIS                      | d    | U00EB                                               | U00CB | af cy de fr fy lb nl sq [Cyr]                                                  |
| WITH HOOK ABOVE                     | d    | U1EBB                                               | U1EBA | vi                                                                             |
| WITH TILDE                          | d    | U1EBD                                               | U1EBC | vi                                                                             |
| WITH DOT ABOVE                      | d    | U0117                                               | U0116 | lt                                                                             |
| WITH DOT BELOW                      | d    | U1EB9                                               | U1EB8 | vi                                                                             |
| WITH OGONEK                         | d    | U0119                                               | U0118 | lt pl                                                                          |
| WITH MACRON                         | d    | U0113                                               | U0112 | lv                                                                             |
| LATIN LETTER F                      | b    | U0066                                               | U0046 | +                                                                              |
| WITH GRAVE                          | d    | —                                                   | —     | [Cyr]                                                                          |
| LATIN LETTER G                      | b    | U0067                                               | U0047 | +                                                                              |
| WITH ACUTE                          | d    | U01F5                                               | U01F4 | [Cyr]                                                                          |
| WITH GRAVE                          | d    | —                                                   | —     | [Cyr]                                                                          |
| WITH BREVE                          | d    | U011F                                               | U011E | tr [Cyr]                                                                       |
| WITH CIRCUMFLEX                     | d    | U011D                                               | U011C | eo                                                                             |
| WITH CARON                          | d    | U01E7                                               | U01E6 | [ar]                                                                           |
| WITH DOT ABOVE                      | d    | U0121                                               | U0120 | mt [Cyr] [ar]                                                                  |
| WITH COMMA BELOW/ABOVE <sup>b</sup> | d    | —                                                   | —     | lv                                                                             |
| WITH HOOK                           | s    | U0260                                               | U0193 |                                                                                |
| LATIN LETTER H                      | b    | U0068                                               | U0048 | +                                                                              |
| WITH CIRCUMFLEX                     | d    | U0125                                               | U0124 | eo                                                                             |

Table D.1 (continued)

| Name                                  | Type | Position for lowercase/uppercase in ISO/IEC 10646-1 |       | Languages used                                                       |
|---------------------------------------|------|-----------------------------------------------------|-------|----------------------------------------------------------------------|
| WITH DOT ABOVE                        | d    | U1E23                                               | U1E22 | [he]                                                                 |
| WITH DOT BELOW                        | d    | U1E25                                               | U1E24 | [Cyr] [ar] [he]                                                      |
| WITH CEDILLA                          | d    | U1E29                                               | U1E28 | [Cyr]                                                                |
| WITH LINE BELOW                       | d    | U1E96                                               | —     | [ar]                                                                 |
| WITH STROKE                           | s    | U0127                                               | U0126 | mt                                                                   |
| LATIN LETTER I                        | b    | U0069                                               | U0049 | +                                                                    |
| LATIN SMALL LETTER DOTLESS I          | b    | U0131                                               |       | tr                                                                   |
| LATIN CAPITAL LETTER I WITH DOT ABOVE | b    |                                                     | U0130 | tr                                                                   |
| LATIN LETTER I WITH ACUTE             | d    | U00ED                                               | U00CD | af ca cs cy da es fo ga gd gl hu is it kl nl pt qal sk vi [Cyr] [ar] |
| WITH GRAVE                            | d    | U00EC                                               | U00CC | cy fur gd it qal vi [Cyr]                                            |
| WITH BREVE                            | d    | U012D                                               | U012C |                                                                      |
| WITH CIRCUMFLEX                       | d    | U00EE                                               | U00CE | af cy fr it kl mo qal ro [Cyr]                                       |
| WITH CARON                            | d    | U01D0                                               | U01CF | [Cyr]                                                                |
| WITH DIAERESIS                        | d    | U00EF                                               | U00CF | af ca cy de fr fy it nl oc [Cyr]                                     |
| WITH HOOK ABOVE                       | d    | U1EC9                                               | U1EC8 | vi                                                                   |
| WITH TILDE                            | d    | U0129                                               | U0128 | kl vi                                                                |
| WITH DOT BELOW                        | d    | U1ECB                                               | U1ECA | vi                                                                   |
| WITH OGONEK                           | d    | U012F                                               | U012E | lt                                                                   |
| WITH MACRON                           | d    | U012B                                               | U012A | lv [Cyr] [ar]                                                        |
| LATIN LETTER J                        | b    | U006A                                               | U004A | +                                                                    |
| WITH ACUTE                            | d    | —                                                   | —     | [Cyr]                                                                |
| WITH CIRCUMFLEX                       | d    | U0135                                               | U0134 | eo                                                                   |
| WITH CARON                            | d    | U01F0                                               | —     | [Cyr]                                                                |
| LATIN LETTER K                        | b    | U006B                                               | U004B | +                                                                    |
| WITH ACUTE                            | d    | U1E31                                               | U1E30 | [Cyr]                                                                |
| WITH GRAVE                            | d    | —                                                   | —     | [Cyr]                                                                |
| WITH CIRCUMFLEX                       | d    | —                                                   | —     | [Cyr]                                                                |
| WITH CARON                            | d    | U01E9                                               | U01E8 | [Cyr]                                                                |
| WITH DOT ABOVE                        | d    | —                                                   | —     | [he]                                                                 |
| WITH DOT BELOW                        | d    | U1E33                                               | U1E32 | [Cyr]                                                                |
| WITH COMMA BELOW <sup>c</sup>         | d    | —                                                   | —     | lv [Cyr]                                                             |
| WITH MACRON                           | d    | —                                                   | —     | [Cyr]                                                                |
| LATIN LETTER K WITH HOOK              | s    | U0199                                               | U0198 | ha                                                                   |
| LATIN SMALL LETTER KRA                | s    | U0138                                               |       | kl                                                                   |
| LATIN LETTER L                        | b    | U006C                                               | U004C | +                                                                    |
| WITH ACUTE                            | d    | U013A                                               | U0139 | [Cyr]                                                                |
| WITH CIRCUMFLEX                       | d    | —                                                   | —     | [Cyr]                                                                |
| WITH CIRCUMFLEX BELOW                 | d    | U1E3D                                               | U1E3C | ve                                                                   |
| WITH CARON                            | d    | U013E                                               | U013D |                                                                      |
| WITH DOT BELOW                        | d    | U1E37                                               | U1E36 |                                                                      |
| WITH COMMA BELOW <sup>d</sup>         | d    | —                                                   | —     | lv [Cyr]                                                             |

Table D.1 (continued)

| Name                           | Type | Position for lowercase/uppercase in ISO/IEC 10646-1 |       | Languages used                                                           |
|--------------------------------|------|-----------------------------------------------------|-------|--------------------------------------------------------------------------|
| WITH STROKE                    | s    | U0142                                               | U0141 | pl                                                                       |
| LATIN LETTER M                 | b    | U006D                                               | U004D | +                                                                        |
| WITH ACUTE                     | d    | U1E3F                                               | U1E3E |                                                                          |
| WITH CIRCUMFLEX                | d    | —                                                   | —     | lb                                                                       |
| WITH DOT ABOVE                 | d    | U1E41                                               | U1E40 |                                                                          |
| WITH DOT BELOW                 | d    | U1E43                                               | U1E42 |                                                                          |
| LATIN LETTER N                 | b    | U006E                                               | U004E | +                                                                        |
| WITH ACUTE                     | d    | U0144                                               | U0143 | pl smi [Cyr]                                                             |
| WITH GRAVE                     | d    | —                                                   | —     | [Cyr]                                                                    |
| WITH BREVE                     | d    | —                                                   | —     | [Cyr]                                                                    |
| WITH CIRCUMFLEX                | d    | —                                                   | —     | lb [Cyr]                                                                 |
| WITH CIRCUMFLEX BELOW          | d    | U1E4B                                               | U1E4A | hz ye                                                                    |
| WITH CARON                     | d    | U0148                                               | U0147 | cs sk                                                                    |
| WITH TILDE                     | d    | U00F1                                               | U00D1 | br es eu gl                                                              |
| WITH DOT ABOVE                 | d    | U1E45                                               | U1E44 | ve [Cyr]                                                                 |
| WITH DOT BELOW                 | d    | U1E47                                               | U1E46 | [Cyr]                                                                    |
| WITH COMMA BELOW <sup>e</sup>  | d    | —                                                   | —     | lv [Cyr]                                                                 |
| WITH MACRON                    | d    | —                                                   | —     | [Cyr]                                                                    |
| PRECEDED BY APOSTROPHE         | d    | U0149                                               | —     |                                                                          |
| FOLLOWED BY APOSTROPHE         | d    | —                                                   | —     | ts                                                                       |
| LATIN LETTER ENG               | s    | U014B                                               | U014A | se                                                                       |
| LATIN LETTER O                 | b    | U006F                                               | U004F | +                                                                        |
| WITH ACUTE                     | d    | U00F3                                               | U00D3 | af ca cs cy da es fo ga gd gl hu is it nl no pl pt qal sk sl ss vi [Cyr] |
| WITH GRAVE                     | d    | U00F2                                               | U00D2 | ca cy fur gd it no pt qal rm vi [Cyr]                                    |
| WITH BREVE                     | d    | U014F                                               | U014E |                                                                          |
| WITH CIRCUMFLEX                | d    | U00F4                                               | U00D4 | af cy de fr fy kl no nso pt qal sk sl tn vi [Cyr]                        |
| WITH CIRCUMFLEX AND ACUTE      | d    | U1ED1                                               | U1ED0 | vi                                                                       |
| WITH CIRCUMFLEX AND GRAVE      | d    | U1ED3                                               | U1ED2 | vi                                                                       |
| WITH CIRCUMFLEX AND HOOK ABOVE | d    | U1ED5                                               | U1ED4 | vi                                                                       |
| WITH CIRCUMFLEX AND TILDE      | d    | U1ED7                                               | U1ED6 | vi                                                                       |
| WITH CIRCUMFLEX AND DOT BELOW  | d    | U1ED9                                               | U1ED8 | vi                                                                       |
| WITH CARON                     | d    | U01D2                                               | U01D1 |                                                                          |
| WITH DIAERESIS                 | d    | U00F6                                               | U00D6 | af cy de et fi fy hu is lb nl rm sv tr [Cyr]                             |
| WITH DIAERESIS AND DOT BELOW   | d    | —                                                   | —     | [Cyr]                                                                    |
| WITH DOUBLE ACUTE              | d    | U0151                                               | U0150 | hu [Cyr]                                                                 |
| WITH HOOK ABOVE                | d    | U1ECF                                               | U1ECE | vi                                                                       |
| WITH TILDE                     | d    | U00F5                                               | U00D5 | et pt vi                                                                 |

Table D.1 (continued)

| Name                        | Type | Position for<br>lowercase/uppercase<br>in ISO/IEC 10646-1 |       | Languages used                                        |
|-----------------------------|------|-----------------------------------------------------------|-------|-------------------------------------------------------|
| WITH DOT BELOW              | d    | U1ECD                                                     | U1ECC | vi                                                    |
| WITH OGONEK                 | d    | U01EB                                                     | U01EA |                                                       |
| WITH MACRON                 | d    | U014D                                                     | U014C | lv [Cyr]                                              |
| WITH HORN                   | d    | U01A1                                                     | U01A0 | vi                                                    |
| WITH HORN AND ACUTE         | d    | U1EDB                                                     | U1EDA | vi                                                    |
| WITH HORN AND GRAVE         | d    | U1EDD                                                     | U1EDC | vi                                                    |
| WITH HORN AND<br>HOOK ABOVE | d    | U1EDF                                                     | U1EDE | vi                                                    |
| WITH HORN AND TILDE         | d    | U1EE1                                                     | U1EE0 | vi                                                    |
| WITH HORN AND DOT BELOW     | d    | U1EE3                                                     | U1EE2 | vi                                                    |
| WITH STROKE                 | s    | U00F8                                                     | U00D8 | da fo is kl no smi                                    |
| WITH STROKE AND ACUTE       | s d  | U01FF                                                     | U01FE |                                                       |
| LATIN LIGATURE OE           | s    | U0153                                                     | U0152 | fr [Cyr]                                              |
| LATIN LETTER P              | b    | U0070                                                     | U0050 | +                                                     |
| WITH ACUTE                  | d    | U1E55                                                     | U1E54 | [Cyr]                                                 |
| WITH GRAVE                  | d    | —                                                         | —     | [Cyr]                                                 |
| WITH DOT ABOVE              | d    | U1E57                                                     | U1E56 | [he]                                                  |
| LATIN LETTER Q              | b    | U0071                                                     | U0051 | +                                                     |
| LATIN LETTER R              | b    | U0072                                                     | U0052 | +                                                     |
| WITH ACUTE                  | d    | U0155                                                     | U0154 | sk                                                    |
| WITH CARON                  | d    | U0159                                                     | U0158 | cs sk                                                 |
| WITH DOT ABOVE              | d    | U1E59                                                     | U1E58 |                                                       |
| WITH DOT BELOW              | d    | U1E5B                                                     | U1E5A |                                                       |
| WITH CEDILLA                | d    | U0157                                                     | U0156 | lv                                                    |
| LATIN LETTER S              | b    | U0073                                                     | U0053 | +                                                     |
| WITH ACUTE                  | d    | U015B                                                     | U015A | pl [he]                                               |
| WITH GRAVE                  | d    | —                                                         | —     | [Cyr] [he]                                            |
| WITH CIRCUMFLEX             | d    | U015D                                                     | U015C | eo [Cyr]                                              |
| WITH CARON                  | d    | U0161                                                     | U0160 | cs et hr lt lv nso sk sl smi sr tn<br>[Cyr] [ar] [he] |
| WITH DOT ABOVE              | d    | U1E61                                                     | U1E60 |                                                       |
| WITH DOT BELOW              | d    | U1E63                                                     | U1E62 | [ar] [he]                                             |
| WITH CEDILLA                | d    | U015F                                                     | U015E | tr                                                    |
| WITH COMMA BELOW            | d    | —                                                         | —     | mo ro [Cyr]                                           |
| LATIN SMALL LETTER SHARP S  | s    | U00DF                                                     |       | de                                                    |
| LATIN LETTER T              | b    | U0074                                                     | U0054 | +                                                     |
| WITH GRAVE                  | d    | —                                                         | —     | [Cyr]                                                 |
| WITH CIRCUMFLEX BELOW       | d    | U1E71                                                     | U1E70 | hz ve                                                 |
| WITH CARON                  | d    | U0165                                                     |       | (cs) (sk) [Cyr]                                       |
| WITH CARON                  | d    |                                                           | U0164 | cs sk [Cyr]                                           |
| WITH DIAERESIS              | d    | U1E97                                                     | —     | [ar]                                                  |
| WITH DOT ABOVE              | d    | U1E6B                                                     | U1E6A |                                                       |
| WITH DOT BELOW              | d    | U1E6D                                                     | U1E6C | [ar] [he]                                             |

Table D.1 (continued)

| Name                          | Type | Position for lowercase/uppercase in ISO/IEC 10646-1 |       | Languages used                                                       |
|-------------------------------|------|-----------------------------------------------------|-------|----------------------------------------------------------------------|
| WITH COMMA BELOW <sup>f</sup> | d    | —                                                   | —     | mo ro [Cyr]                                                          |
| WITH LINE BELOW               | d    | U1E6F                                               | U1E6E | [ar]                                                                 |
| FOLLOWED BY APOSTROPHE        | d    | —                                                   |       | cs sk                                                                |
| WITH STROKE                   | s    | U0167                                               | U0166 | se                                                                   |
| LATIN LETTER U                | b    | U0075                                               | U0055 | +                                                                    |
| WITH ACUTE                    | d    | U00FA                                               | U00DA | af ca cs cy da es fo fy ga gl hu is it kl nl pt qal sk vi [Cyr] [ar] |
| WITH GRAVE                    | d    | U00F9                                               | U00D9 | br cy fr fur gd it qal vi [Cyr]                                      |
| WITH BREVE                    | d    | U016D                                               | U016C | eo [Cyr]                                                             |
| WITH CIRCUMFLEX               | d    | U00FB                                               | U00DB | af cy fr fy kl qal tr [Cyr]                                          |
| WITH CARON                    | d    | U01D4                                               | U01D3 |                                                                      |
| WITH RING ABOVE               | d    | U016F                                               | U016E | cs [Cyr]                                                             |
| WITH DIAERESIS                | d    | U00FC                                               | U00DC | br ca cy de es et fr fy gl hu lb nl pt rm tr [Cyr]                   |
| WITH DIAERESIS AND DOT BELOW  | d    | —                                                   | —     | [Cyr]                                                                |
| WITH DOUBLE ACUTE             | d    | U0171                                               | U0170 | hu [Cyr]                                                             |
| WITH HOOK ABOVE               | d    | U1EE7                                               | U1EE6 | vi                                                                   |
| WITH TILDE                    | d    | U0169                                               | U0168 | kl vi                                                                |
| WITH DOT ABOVE                | d    |                                                     | —     | [Cyr]                                                                |
| WITH DOT BELOW                | d    | U1EE5                                               | U1EE4 | vi                                                                   |
| WITH OGONEK                   | d    | U0173                                               | U0172 | lt                                                                   |
| WITH MACRON                   | d    | U016B                                               | U016A | lt lv [Cyr] [ar]                                                     |
| WITH MACRON AND DOT BELOW     | d    | —                                                   | —     | [Cyr]                                                                |
| WITH HORN                     | d    | U01B0                                               | U01AF | vi                                                                   |
| WITH HORN AND ACUTE           | d    | U1EE9                                               | U1EE8 | vi                                                                   |
| WITH HORN AND GRAVE           | d    | U1EEB                                               | U1EEA | vi                                                                   |
| WITH HORN AND HOOK ABOVE      | d    | U1EED                                               | U1EEC | vi                                                                   |
| WITH HORN AND TILDE           | d    | U1EEF                                               | U1EEE | vi                                                                   |
| WITH HORN AND DOT BELOW       | d    | U1EF1                                               | U1EF0 | vi                                                                   |
| LATIN LETTER V                | b    | U0076                                               | U0056 | +                                                                    |
| WITH TILDE                    | d    | U1E7D                                               | U1E7C |                                                                      |
| WITH DOT BELOW                | d    | U1E7F                                               | U1E7E |                                                                      |
| LATIN LETTER W                | b    | U0077                                               | U0057 | +                                                                    |
| WITH ACUTE                    | d    | U1E83                                               | U1E82 | cy                                                                   |
| WITH GRAVE                    | d    | U1E81                                               | U1E80 | cy                                                                   |
| WITH CIRCUMFLEX               | d    | U0175                                               | U0174 | cy [he]                                                              |
| WITH DIAERESIS                | d    | U1E85                                               | U1E84 | cy                                                                   |
| WITH DOT ABOVE                | d    | U1E87                                               | U1E86 | [he]                                                                 |
| WITH DOT BELOW                | d    | U1E89                                               | U1E88 |                                                                      |



Table D.1 (continued)

| Name                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        | Type | Position for lowercase/uppercase in ISO/IEC 10646-1 |       | Languages used                 |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|-----------------------------------------------------|-------|--------------------------------|
| LATIN LETTER X                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | b    | U0078                                               | U0058 | +                              |
| WITH DIAERESIS                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | d    | U1E8D                                               | U1E8C |                                |
| WITH DOT ABOVE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | d    | U1E8B                                               | U1E8A |                                |
| LATIN LETTER Y                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | b    | U0079                                               | U0059 | +                              |
| WITH ACUTE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | d    | U00FD                                               | U00DD | af cs cy is fo sk vi           |
| WITH GRAVE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | d    | U1EF3                                               | U1EF2 | cy vi [Cyr] [ar]               |
| WITH CIRCUMFLEX                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | d    | U0177                                               | U0176 | cy                             |
| WITH DIAERESIS                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | d    | U00FF                                               | U0178 | cy fr [Cyr]                    |
| WITH HOOK ABOVE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | d    | U1EF7                                               | U1EF6 | vi                             |
| WITH TILDE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | d    | U1EF9                                               | U1EF8 | vi                             |
| WITH DOT ABOVE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | d    | U1E8F                                               | U1E8E |                                |
| WITH DOT BELOW                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | d    | U1EF5                                               | U1EF4 | vi                             |
| WITH MACRON                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | d    | —                                                   | —     | [Cyr]                          |
| LATIN LETTER Z                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | b    | U007A                                               | U005A | +                              |
| WITH ACUTE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | d    | U017A                                               | U0179 | pl [Cyr]                       |
| WITH BREVE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | d    | —                                                   | —     | [Cyr]                          |
| WITH CIRCUMFLEX                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | d    | U1E91                                               | U1E90 | [Cyr]                          |
| WITH CARON                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  | d    | U017E                                               | U017D | cs hr lt lv sk sl smi sr [Cyr] |
| WITH CARON AND CEDILLA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      | d    | —                                                   | —     | [Cyr]                          |
| WITH DIAERESIS                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | d    | —                                                   | —     | [Cyr]                          |
| WITH DOT ABOVE                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | d    | U017C                                               | U017B | mt pl                          |
| WITH DOT BELOW                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              | d    | U1E93                                               | U1E92 | [ar]                           |
| WITH CEDILLA                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                | d    | —                                                   | —     | [Cyr]                          |
| WITH MACRON                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 | d    | —                                                   | —     | [Cyr]                          |
| LATIN LETTER THORN                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          | b    | U00FE                                               | U00DE | is                             |
| <p><sup>a</sup> LATIN CAPITAL LETTER D WITH STROKE and LATIN CAPITAL LETTER ETH have identical form.</p> <p><sup>b</sup> The character is called LATIN LETTER G WITH CEDILLA in ISO/IEC 10646-1, positions U0123 and U0122.</p> <p><sup>c</sup> The character is called LATIN LETTER K WITH CEDILLA in ISO/IEC 10646-1, positions U0137 and U0136.</p> <p><sup>d</sup> The character is called LATIN LETTER L WITH CEDILLA in ISO/IEC 10646-1, positions U013C and U013B.</p> <p><sup>e</sup> The character is called LATIN LETTER N WITH CEDILLA in ISO/IEC 10646-1, positions U0146 and U0145.</p> <p><sup>f</sup> The character is called LATIN LETTER T WITH CEDILLA in ISO/IEC 10646-1, positions U0163 and U0162.</p> |      |                                                     |       |                                |

## Annex E (informative)

### Languages using the Latin alphabet

[Table E.1](#) lists languages using the Latin alphabet. The alpha-2 symbols are taken from ISO 639-1<sup>[7]</sup>; the alpha-3 symbols are taken from ISO 639-2<sup>[8]</sup> (the terminology code is used when this differs from the bibliographic code). For some languages that are not included in ISO 639-1 or ISO 639-2, ad hoc symbols starting with “qa” have been assigned for reference.

[Table E.2](#) lists alpha-2 and alpha-3 symbols and English names of languages in alphabetical order by alpha-2 and alpha-3 symbols.

**Table E.1 — Alphabetical list by English names of language**

| alpha-2 code | alpha-3 code | English name of language      | French name of language |
|--------------|--------------|-------------------------------|-------------------------|
| af           | afr          | Afrikaans                     | afrikaans               |
| ak           | aka          | Akan                          | akan                    |
| sq           | sqi          | Albanian                      | albanais                |
| ay           | aym          | Aymara                        | aymara                  |
| az           | aze          | Azerbaijani                   | azéri; azerbaïdjanaïs   |
| eu           | eus          | Basque                        | basque                  |
| bi           | bis          | Bislama                       | bichlamar; bislama      |
| br           | bre          | Breton                        | breton                  |
| ca           | cat          | Catalan                       | catalan                 |
| ch           | cha          | Chamorro                      | chamorro                |
| kw           | cor          | Cornish                       | cornique                |
| co           | cos          | Corsican                      | corse                   |
| hr           | hrv          | Croatian                      | croate                  |
| cs           | ces          | Czech                         | tchèque                 |
| da           | dan          | Danish                        | danois                  |
| nl           | nld          | Dutch                         | néerlandais             |
| —            | efi          | Efik                          | efik                    |
| en           | eng          | English                       | anglais                 |
| eo           | epo          | Esperanto                     | espéranto               |
| et           | est          | Estonian                      | estonien                |
| ee           | ewe          | Ewe                           | éwé                     |
| fo           | fao          | Faroese; Faeroese             | féroïen                 |
| fj           | fij          | Fiji; Fijian                  | fidji; fidjien          |
| fi           | fin          | Finnish                       | finnois                 |
| fr           | fra          | French                        | français                |
| fy           | fry          | Frisian                       | frison                  |
| —            | fur          | Friulian                      | frioulan                |
| ff           | ful          | Fulah; Fula; Fulani; Fulfulde | peol                    |
| gl           | glg          | Galician; Gallegan            | galicien                |
| de           | deu          | German                        | allemand                |

Table E.1 (continued)

| alpha-2 code | alpha-3 code | English name of language | French name of language                           |
|--------------|--------------|--------------------------|---------------------------------------------------|
| –            | gil          | Gilbertese               | kiribati                                          |
| kl           | kal          | Greenlandic; Kalaallisut | groenlandais                                      |
| ha           | hau          | Hausa                    | haoussa                                           |
| –            | haw          | Hawaiian                 | hawaïen                                           |
| hz           | her          | Herero                   | herero                                            |
| ho           | hmo          | Hiri Motu                | hiri motu; hiri motou                             |
| hu           | hun          | Hungarian                | hongrois                                          |
| is           | isl          | Icelandic                | islandais                                         |
| –            | ibo          | Igbo                     | igbo                                              |
| id           | ind          | Indonesian               | indonésien                                        |
| ia           | ina          | Interlingua              | interlingua                                       |
| ie           | ile          | Interlingue              | interlingue                                       |
| iu           | iku          | Inuktitut                | inuktitut                                         |
| ik           | ipk          | Inupiaq; Inupiak         | inupiaq                                           |
| ga           | gle          | Irish; Irish Gaelic      | irlandais; gaélique d'Irlande; gaélique irlandais |
| it           | ita          | Italian                  | italien                                           |
| jv           | jaw          | Javanese                 | javanais                                          |
| kr           | kau          | Kanuri                   | kanouri                                           |
| ki           | kik          | Kikuyu; Gikuyu           | kikuyu                                            |
| rw           | kin          | Kinyarwanda; Rwanda      | kinyarwanda; rwanda                               |
| rn           | run          | Kirundi; Rundi           | kirundi; rundi                                    |
| kg           | kon          | Kongo; Kikongo           | kongo                                             |
| ku           | kur          | Kurdish                  | kurde                                             |
| kj           | kua          | Kwanyama                 | kwanyama                                          |
| –            | qal          | Ladin                    | ladin                                             |
| –            | lad          | Ladino                   | ladino; judéo-espagnol                            |
| la           | lat          | Latin                    | latin                                             |
| lv           | lav          | Latvian; Lettish         | letton; lettonien                                 |
| ln           | lin          | Lingala                  | lingala                                           |
| lt           | lit          | Lithuanian               | lituanien                                         |
| lu           | lub          | Luba; Luba-Katanga       | tchiluba; luba-katanga                            |
| lg           | lug          | Luganda; Ganda           | louganda; ganda                                   |
| lb           | ltz          | Luxembourgish            | luxembourgeois                                    |
| mg           | mlg          | Malagasy                 | malagache                                         |
| ms           | msa          | Malay                    | malais                                            |
| mt           | mlt          | Maltese                  | maltais                                           |
| –            | man          | Mandingo                 | mandingue                                         |
| gv           | glv          | Manx Gaelic              | gaélique de l'Île de Man; manx                    |
| mi           | mri          | Maori                    | maori                                             |
| na           | nau          | Nauru; Nauruan           | nauruen; nauri                                    |
| nd           | nde          | Ndebele, North           | ndébèle du Nord                                   |
| nr           | nbl          | Ndebele, South           | ndébèle du Sud                                    |
| ng           | ndo          | Ndonga                   | ndonga                                            |
| no           | nor          | Norwegian                | norvégien                                         |

Table E.1 (continued)

| alpha-2 code | alpha-3 code | English name of language          | French name of language                 |
|--------------|--------------|-----------------------------------|-----------------------------------------|
| oc           | oci          | Occitan; Provençal                | occitan; provençal                      |
| oj           | oji          | Ojibway                           | ojibwa                                  |
| pl           | pol          | Polish                            | polonais                                |
| pt           | por          | Portuguese                        | portugais                               |
| qu           | que          | Quechua                           | quechua; quichua                        |
| rm           | roh          | Raeto-Romance                     | rhéto-roman                             |
| ro           | ron          | Romanian<br>Moldavian<br>Moldovan | roumain<br>moldave                      |
| –            | smi          | Sami                              | sami                                    |
| se           | sme          | Sami, Northern                    | sami du Nord                            |
| sm           | smo          | Samoan                            | samoan                                  |
| sg           | sag          | Sango; Sangho                     | sango                                   |
| sc           | srd          | Sardinian                         | sarde                                   |
| gd           | gla          | Scottish Gaelic                   | gaélique de l'Écosse; gaélique écossais |
| sr           | srp          | Serbian                           | serbe                                   |
| st           | sot          | Sesotho                           | sesotho; sotho du Sud                   |
| tn           | tsn          | Setswana; Tswana                  | setchwana; tswana                       |
| sn           | sna          | Shona; chiShona                   | shona                                   |
| ss           | ssw          | Siswati; Swati                    | siswati; swati; swazi                   |
| sk           | slk          | Slovak                            | slovaque                                |
| sl           | slv          | Slovenian                         | slovène                                 |
| so           | som          | Somali                            | somali                                  |
| –            | wen          | Sorbian                           | sorabe                                  |
| –            | nso          | Sotho, Northern                   | sotho du Nord                           |
| es           | spa          | Spanish; Castilian                | espagnol, castillan                     |
| su           | sun          | Sundanese                         | soundanais                              |
| sw           | swa          | Swahili; Kiswahili                | swahili                                 |
| sv           | swe          | Swedish                           | suédois                                 |
| tl           | tgl          | Tagalog                           | tagal; tagalog                          |
| ty           | tah          | Tahitian                          | tahitien                                |
| to           | ton          | Tonga; Tongan                     | tonga                                   |
| ts           | tso          | Tsonga                            | tsonga                                  |
| tr           | tur          | Turkish                           | turc                                    |
| –            | tvl          | Tuvalu                            | tuvalu                                  |
| tw           | twi          | Twí; Tshi                         | twi; tchi                               |
| ug           | uig          | Uyghur                            | ouïgour                                 |
| uz           | uzb          | Uzbek                             | ouzbek                                  |
| ve           | ven          | Venda                             | venda                                   |
| vi           | vie          | Vietnamese                        | vietnamien                              |
| vo           | vol          | Volapük                           | volapük                                 |
| cy           | cym          | Welsh                             | gallois                                 |
| wo           | wol          | Wolof                             | wolof                                   |
| xh           | xho          | Xhosa                             | xhosa                                   |

Table E.1 (continued)

| alpha-2 code | alpha-3 code | English name of language | French name of language |
|--------------|--------------|--------------------------|-------------------------|
| yo           | yor          | Yoruba                   | yorouba; yoruba         |
| za           | zha          | Zhuang; Chwang; Chuang   | zhuang; jouang          |
| zu           | zul          | Zulu                     | zoulou                  |

Table E.2 — Alphabetical list by language symbols

| alpha-2 and -3 code | English name of language      |
|---------------------|-------------------------------|
| af afr              | Afrikaans                     |
| ak aka              | Akan                          |
| ay aym              | Aymara                        |
| az aze              | Azerbaijani                   |
| bi bis              | Bislama                       |
| br bre              | Breton                        |
| ca cat              | Catalan                       |
| ces                 | Czech                         |
| ch cha              | Chamorro                      |
| co                  | Corsican                      |
| cor                 | Cornish                       |
| cos                 | Corsican                      |
| cs                  | Czech                         |
| cy cym              | Welsh                         |
| da dan              | Danish                        |
| de deu              | German                        |
| ee                  | Ewe                           |
| efi                 | Efik                          |
| en eng              | English                       |
| eo epo              | Esperanto                     |
| es                  | Spanish; Castilian            |
| est et              | Estonian                      |
| eu eus              | Basque                        |
| ewe                 | Ewe                           |
| fao                 | Faroese; Faeroese             |
| ff                  | Fulah; Fula; Fulani; Fulfulde |
| fi                  | Finnish                       |
| fij                 | Fiji; Fijian                  |
| fin                 | Finnish                       |
| fj                  | Fiji; Fijian                  |
| fo                  | Faroese; Faeroese             |
| fr fra              | French                        |
| fry                 | Frisian                       |
| ful                 | Fulah; Fula; Fulani; Fulfulde |
| fur                 | Friulian                      |
| fy                  | Frisian                       |
| ga                  | Irish; Irish Gaelic           |
| gd                  | Scottish Gaelic               |

Table E.2 (continued)

| alpha-2 and -3 code | English name of language |
|---------------------|--------------------------|
| gil                 | Gilbertese               |
| gl                  | Galician; Gallegan       |
| gla                 | Scottish Gaelic          |
| gle                 | Irish; Irish Gaelic      |
| glg                 | Galician; Gallegan       |
| glv gv              | Manx Gaelic              |
| ha hau              | Hausa                    |
| haw                 | Hawaiian                 |
| her                 | Herero                   |
| hmo ho              | Hiri Motu                |
| hr hrv              | Croatian                 |
| hu hun              | Hungarian                |
| hz                  | Herero                   |
| ia                  | Interlingua              |
| ibo                 | Igbo                     |
| id                  | Indonesian               |
| ie                  | Interlingue              |
| ik                  | Inupiaq; Inupiak         |
| iku                 | Inuktitut                |
| ile                 | Interlingue              |
| ina                 | Interlingua              |
| ind                 | Indonesian               |
| ipk                 | Inupiaq; Inupiak         |
| is isl              | Icelandic                |
| it ita              | Italian                  |
| iu                  | Inuktitut                |
| jaw jv              | Javanese                 |
| kal                 | Greenlandic              |
| kau                 | Kanuri                   |
| kg                  | Kongo; Kikongo           |
| ki kik              | Kikuyu; Gikuyu           |
| kin                 | Kinyarwanda; Rwanda      |
| kj                  | Kwanyama                 |
| kl                  | Greenlandic              |
| kon                 | Kongo; Kikongo           |
| kr                  | Kanuri                   |
| ku                  | Kurdish                  |
| kua                 | Kwanyama                 |
| kur                 | Kurdish                  |
| kw                  | Cornish                  |
| la                  | Latin                    |
| lad                 | Ladino                   |
| lat                 | Latin                    |
| lav                 | Latvian; Lettish         |

Table E.2 (continued)

| alpha-2 and -3 code | English name of language |
|---------------------|--------------------------|
| lb                  | Luxemburgish             |
| lg                  | Luganda; Ganda           |
| lin                 | Lingala                  |
| lit                 | Lithuanian               |
| ln                  | Lingala                  |
| lt                  | Lithuanian               |
| ltz                 | Luxemburgish             |
| lu lub              | Luba; Luba-Katanga       |
| lug                 | Luganda; Ganda           |
| lv                  | Latvian; Lettish         |
| man                 | Mandingo                 |
| mg                  | Malagasy                 |
| mi                  | Maori                    |
| mlg                 | Malagasy                 |
| mlt                 | Maltese                  |
| mri                 | Maori                    |
| ms msa              | Malay                    |
| mt                  | Maltese                  |
| na nau              | Nauru; Nauruan           |
| nbl                 | Ndebele, South           |
| nd nde              | Ndebele, North           |
| ndo ng              | Ndonga                   |
| nl nld              | Dutch                    |
| no nor              | Norwegian                |
| nr                  | Ndebele, South           |
| nso                 | Sotho, Northern          |
| oc oci              | Occitan; Provençal       |
| oj oji              | Ojibway                  |
| pl pol              | Polish                   |
| por pt              | Portuguese               |
| qal                 | Ladin                    |
| qu que              | Quechua                  |
| rm                  | Raeto-Romance            |
| rn                  | Kirundi; Rundi           |
| ro                  | Romanian, Moldavian      |
| roh                 | Raeto-Romance            |
| ron                 | Romanian                 |
| run                 | Kirundi; Rundi           |
| rw                  | Kinyarwanda; Rwanda      |
| sag                 | Sango; Sangho            |
| sc                  | Sardinian                |
| se                  | Sami, Northern           |
| sg                  | Sango; Sangho            |
| sk                  | Slovak                   |

Table E.2 (continued)

| alpha-2 and -3 code | English name of language |
|---------------------|--------------------------|
| sl                  | Slovenian                |
| slk                 | Slovak                   |
| slv                 | Slovenian                |
| sm                  | Samoan                   |
| sme                 | Sami, Northern           |
| smi                 | Sami                     |
| smo                 | Samoan                   |
| sn sna              | Shona; chiShona          |
| so som              | Somali                   |
| sot                 | Sesotho                  |
| spa                 | Spanish; Castilian       |
| sq sqi              | Albanian                 |
| sr                  | Serbian                  |
| srd                 | Sardinian                |
| srp                 | Serbian                  |
| ss ssw              | Siswati; Swati           |
| st                  | Sesotho                  |
| su sun              | Sundanese                |
| sv                  | Swedish                  |
| sw swa              | Swahili; Kiswahili       |
| swe                 | Swedish                  |
| tah                 | Tahitian                 |
| tgl tl              | Tagalog                  |
| tn                  | Setswana; Tswana         |
| to ton              | Tonga; Tongan            |
| tr                  | Turkish                  |
| ts                  | Tsonga                   |
| tsn                 | Setswana; Tswana         |
| tso                 | Tsonga                   |
| tur                 | Turkish                  |
| tv                  | Tuvalu                   |
| tw twi              | Twi; Tshi                |
| ty                  | Tahitian                 |
| ug uig              | Uyghur                   |
| uz uzb              | Uzbek                    |
| ve ven              | Venda                    |
| vi vie              | Vietnamese               |
| vo vol              | Volapük                  |
| wen                 | Sorbian                  |
| wo wol              | Wolof                    |
| xh xho              | Xhosa                    |
| yo yor              | Yoruba                   |
| za zha              | Zhuang; Chwang; Chuang   |
| zu zul              | Zulu                     |



## Annex F (informative)

### Alphabetical sequences and character repertoires

Most languages deviate from the rules that are laid down in this document. [Table F.1](#) formalizes rules given in national standards and other documents that will have to be followed in addition to, or instead of, the rules given in this document to produce alphabetically ordered lists in these languages. Where no reference is given, the information is derived from dictionaries, etc.

Dictionaries and other alphabetically arranged lists in many languages may be based on different ordering rules and even different characters from those given below or those specified in national standards.

Deviations are grouped as follows:

- ① deviations as to first-ordering-level values; see [Clause 5](#);
- ② deviations as to other ordering rules; see [Clauses 6 to 8](#);
- ③ letters with diacritical marks and special letters that are native to the language; within parentheses, items that are used in loan words and names only; see [5.3](#) and [6.2](#);
- ④ basic Latin letters that are not native to the language (but can be used in, for example, names); within parentheses, items that are used in loan words; see [5.2](#), item b).

Lowercase letters are used to represent both lowercase and uppercase letters. See [Table D.1](#) for the correspondence between lowercase and uppercase letters.

Basic Latin letters are shown in their typographical form. Most other letters are shown in their typographical form followed by the position in ISO/IEC 10646-1. Some letters are replaced by #1, #2, etc. followed by the position and character name in ISO/IEC 10646-1. U???? indicates that the character is not included in ISO/IEC 10646-1.

The following notation is used:

- + sequence of items following a letter;
- = items that are treated as equivalent;
- deviations from the rules of character equivalence given in [5.3](#).

[Table F.2](#) lists the character repertoires of some internationally standardized transliteration systems in Latin script.

NOTE 1 The languages that are included in [Tables F.1](#) and [F.2](#) are those for which information has been made available.

NOTE 2 In [Table F.1](#) “Variants” are, for example, geographical, temporal or linguistic variants. For references, see the Bibliography.

Table F.1 — Alphabetical sequence and character repertoires

| Language                     | Variants;<br>References | Deviations and other information on alphabet<br>(① first level; ② other levels; ③ native letters; ④ non-native)                                                                                             |
|------------------------------|-------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Afrikaans</b><br>af / afr |                         | ① –<br>② –<br>③ á U00E1 é U00E9 è U00E8 ê U00EA ë U00EB í U00ED<br>î U00EE ï U00EF ó U00F3 ô U00F4 ö U00F6 ú U00FA<br>û U00FB ý U00FD (à) U00E0 (ä) U00E4 (ç) U00E7<br>(i) U00EC (ò) U00F2 (ù) U00F9<br>④ þ |
| <b>Albanian</b><br>sq / sqi  |                         | ① c + ç d + dh e + ë g + gj l + ll n + nj r + rr<br>s + sh t + th x + xh z + zh<br>② –<br>③ ç U00E7 ë U00EB<br>④ þ                                                                                          |
| <b>Basque</b><br>eu / eus    |                         | ① l + ll n + ñ t + tt<br>② –<br>③ ñ U00F1 (ç) U00E7 (ü) U00FC<br>④ þ                                                                                                                                        |
| <b>Breton</b><br>br / bre    |                         | ① c + ch + c'h n + ñ<br>② –<br>③ â U00E2 ê U00EA ñ U00F1 ù U00F9 ü U00FC<br>④ þ (c) (q) (x)                                                                                                                 |
| <b>Catalan</b><br>ca / cat   |                         | ① l + ll<br>l.l or l:l is ordered between lk and lm.<br>② –<br>③ –<br>④ á U00E1 à U00E0 ç U00E7 é U00E9 è U00E8 í U00ED<br>î U00EF ó U00F3 ò U00F2 ú U00FA ü U00FC<br>þ                                     |
| <b>Cornish</b><br>kw / cor   |                         | ① –<br>② –<br>③ –<br>④ þ (x) (z)                                                                                                                                                                            |
| <b>Corsican</b><br>co / cos  |                         | ① g + ghj z + chj<br>② –<br>③ à U00E0 è U00E8 ì U00EC ò U00F2 ù U00F9<br>④ w x y þ<br>j is used only in the polygraphs ghj and chj                                                                          |
| <b>Croatian</b><br>hr / hrv  |                         | ① c + č + ć d + đž + đ l + lj n + nj s + š z + ž<br>② –<br>③ č U010D ć U0107 đ U0111 š U0161 ž U017E<br>④ þ (q) (w) (x) (y)                                                                                 |

Table F.1 (continued)

| Language                     | Variants;<br>References                              | Deviations and other information on alphabet<br>(① first level; ② other levels; ③ native letters; ④ non-native)                                                                                                                                                                                                                                                                                                                                                          |
|------------------------------|------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Czech</b><br>cs / ces     |                                                      | <p>① c + č h + ch r + ř s + š z + ž</p> <p>② –</p> <p>③ á U00E1 č U010D d' U???? Ď U010E é U00E9 ě U011B<br/>í U00ED ě U0148 ó U00F3 ř U0159 š U0161 ť U????<br/>Ť U0164 ú U00FA ů U016F ý U00FD ž U017E<br/>d' (lowercase) corresponds to Ď (uppercase) and is sometimes represented as U010F LATIN SMALL LETTER D WITH CARON.<br/>ť (lowercase) corresponds to Ť (uppercase) and is sometimes represented as U0165 LATIN SMALL LETTER T WITH CARON.</p> <p>④ ě (w)</p> |
| <b>Danish</b><br>da / dan    | DS 377 and Dansk sprognævn (Danish Language Council) | <p>① z + æ + ø + å ü = y ŷ = y ä = æ ö = ø ø = ø<br/>þ → th<br/>aa → å (except when the two as belong to separate syllables)<br/>Word-by-word ordering is given priority.<br/>Numbers shall be transcribed to words before ordering.</p> <p>② Uppercase letter before lowercase letter.</p> <p>③ á U00E1 å U00E5 æ U00E6 é U00E9 í U00ED<br/>ó U00F3 ø U00F8 ú U00FA (ä) U00E4 (ö) U00F6<br/>(ø) U0151 (ü) U00FC (ŷ) U0171</p> <p>④ þ</p>                                |
| <b>Dutch</b><br>nl / nld     |                                                      | <p>① –<br/>The digraph ij may be treated as equivalent to y, especially in lists of names.</p> <p>② –</p> <p>③ á U00E1 à U00E0 ä U00E4 é U00E9 è U00E8<br/>ê U00EA ë U00EB í U00ED ì U00EC ï U00EF<br/>ó U00F3 ò U00F2 ö U00F6 ú U00FA ù U00F9 ü U00FC<br/>The digraph ij/IJ was previously regarded as a ligature. The initial uppercase form is IJ (not "Ij").</p> <p>④ þ</p>                                                                                          |
| <b>English</b><br>en / eng   | BS 1749                                              | <p>① þ → th<br/>Letter-by-letter and word-by-word ordering have equal status.</p> <p>② –</p> <p>③ (à) U00E0 (æ) U00E6 (ç) U00E7 (é) U00E9 (è) U00E8<br/>(ë) U00EB (ï) U00EF (ñ) U00F1 (ö) U00F6 (ø) U00F4<br/>(œ) U0153</p> <p>④ þ</p>                                                                                                                                                                                                                                   |
| <b>Esperanto</b><br>eo / epo |                                                      | <p>① c + ĉ g + ĝ h + ĥ j + ĵ s + ŝ u + ŭ</p> <p>② –</p> <p>③ ĉ U0109 ĝ U011D ĥ U0125 ĵ U0135 ŝ U015D ŭ U016D</p> <p>④ ĵ (q) (w) (x) (y)</p>                                                                                                                                                                                                                                                                                                                              |
| <b>Estonian</b><br>et / est  |                                                      | <p>① s + š (z) + õ + ä + ö + ü</p> <p>② –</p> <p>③ ä U00E4 ö U00F6 õ U00F5 š U0161 ü U00FC</p> <p>④ z þ (c) (q) (w) (x) (y)</p>                                                                                                                                                                                                                                                                                                                                          |

Table F.1 (continued)

| Language                                  | Variants;<br>References | Deviations and other information on alphabet<br>(① first level; ② other levels; ③ native letters; ④ non-native)                                                                                                              |
|-------------------------------------------|-------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Faroese;<br/>Faeroese</b><br>fo / fao  |                         | ① a + á d + ð i + í o + ó u + ú y + ý z + æ + ø<br>② –<br>③ á U00E1 æ U00E6 ð U00F0 í U00ED ó U00F3<br>ø U00F8 ú U00FA ý U00FD<br>④ þ                                                                                        |
| <b>Finnish</b><br>fi / fin                | SFS 4600                | ① š = s ž = z z + å + ä + ö w = v ü = y ø = ö<br>þ → th<br>② –<br>③ ä U00E4 ö U00F6 (å) U00E5 (š) U0161 (ž) U017E<br>④ þ (b) (c) (f) (q) (w)                                                                                 |
| <b>French</b><br>fr / fra                 |                         | ① –<br>② Diacritics from right to left for each word.<br>③ à U00E0 â U00E2 æ U00E6 ç U00E7 é U00E9<br>è U00E8 ê U00EA ë U00EB î U00EE ï U00EF<br>ô U00F4 œ U0153 ù U00F9 û U00FB ü U00FC<br>ÿ U00FF<br>④ þ                   |
| <b>Frisian</b><br>fy / fry                |                         | ① ij = y<br>② –<br>③ à U00E0 â U00E2 ä U00E4 é U00E9 ê U00EA<br>ë U00EB ï U00EF ô U00F4 ö U00F6 ú U00FA<br>û U00FB ü U00FC<br>④ þ                                                                                            |
| <b>Friulian</b><br>– / fur                |                         | ① a + à + á + â e + è i + ì o + ò u + ù<br>② –<br>③ á U00E1 à U00E0 â U00E2 è U00E8 ì U00EC ò U00F2<br>ù U00F9<br>④ þ                                                                                                        |
| <b>Galician;<br/>Gallegan</b><br>gl / glg |                         | ① n + ñ<br>② –<br>③ á U00E1 é U00E9 í U00ED ñ U00F1 ó U00F3 ú U00FA<br>ü U00F<br>④ þ                                                                                                                                         |
| <b>German</b><br>de / deu                 | DIN 5007<br>ON A 2725   | ① þ → th<br>Word-by-word ordering specified.<br>② Ordering according to diacritical marks: “Umlaut” first.<br>③ à U00E0 â U00E2 ä U00E4 é U00E9 è U00E8<br>ê U00EA ë U00EB ï U00EF ô U00F4 ö U00F6<br>ß U00DF ü U00FC<br>④ þ |

Table F.1 (continued)

| Language                                        | Variants;<br>References                                  | Deviations and other information on alphabet<br>(① first level; ② other levels; ③ native letters; ④ non-native)                                                                                                                                                                                                                                                                                                                                       |
|-------------------------------------------------|----------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Greenlandic;<br/>Kalaallisut</b><br>kl / kal |                                                          | ① z + æ + ø + å      K = q<br>② –<br>③ á U00E1      â U00E2      ð U00E5      ã U00E3      æ U00E6<br>é U00E9      í U00ED      î U00EE      ï U0129      k U0138<br>ô U00F4      ø U00F8      ú U00FA      û U00FB      ü U0169<br>K (U0138 LATIN SMALL LETTER KRA) has been replaced in the orthography by the letter q. The uppercase letter corresponding to LATIN SMALL LETTER KRA is K' (LATIN CAPITAL LETTER K FOLLOWED BY APOSTROPHE).<br>④ þ |
| <b>Hausa</b><br>ha / hau                        |                                                          | ① b + #1      d + #2      k + #3<br>② –<br>③ #1 U0253 B WITH HOOK      #2 U0257 D WITH HOOK<br>#3 U0199 K WITH HOOK<br>④ j      q      v      x      þ                                                                                                                                                                                                                                                                                                |
| <b>Herero</b><br>hz / her                       |                                                          | ① a + e + i + o + u +<br>h + k + m + n + #2 + p + r + s + t + #3 + v + w + y + z +<br>b + d + f + g + j + l +<br>c + q + x +<br>mb + nd + n#1 + ng + ny + ndj + tj<br>② –<br>③ #1 U1E13 D WITH CIRCUMFLEX BELOW<br>#2 U1E4B N WITH CIRCUMFLEX BELOW<br>#3 U1E71 T WITH CIRCUMFLEX BELOW<br>④ c      q      x      þ                                                                                                                                   |
| <b>Hungarian</b><br>hu / hun                    |                                                          | ① c + cs      d + dz + dzs      g + gy      l + ly      n + ny      o + ö      s + sz<br>t + ty      u + ü      z + zs      ő = ö      ű = ű<br>② –<br>③ á U00E1      é U00E9      í U00ED      ó U00F3      ö U00F6      ő U0151<br>ú U00FA      ü U00FC      ű U0171<br>④ þ      (q)      (w)      (x)                                                                                                                                              |
| <b>Icelandic</b><br>is / isl                    | Íslensk<br>málefnd<br>(Icelandic<br>Language<br>Council) | ① a + á      d + ð      e + é      i + í      o + ó      u + ú      y + ý<br>þ + æ + ö + (ǿ)      ä = æ      ø = ö<br>② –<br>③ á U00E1      æ U00E6      ð U00F0      é U00E9      í U00ED<br>ó U00F3      ö U00F6      ú U00FA      ý U00FD      (ǿ) U00E4<br>(ǿ) U00E5      (ø) U00F8<br>④ c      q      w      z                                                                                                                                   |
| <b>Irish;<br/>Irish Gaelic</b><br>ga / gle      |                                                          | ① –<br>② –<br>③ á U00E1      é U00E9      í U00ED      ó U00F3      ú U00FA<br>④ þ      (k)<br>Irish also uses a Gaelic variant of the Latin script. This script has not been included in the character set of this document.                                                                                                                                                                                                                         |

Table F.1 (continued)

| Language                                | Variants;<br>References | Deviations and other information on alphabet<br>(① first level; ② other levels; ③ native letters; ④ non-native)                                                                                                                                                                                                                                                                      |
|-----------------------------------------|-------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Italian</b><br>it / ita              |                         | ① –<br>② –<br>③ à U00E0 é U00E9 è U00E8 í U00ED ì U00EC î U00EE<br>ï U00EF ó U00F3 ò U00F2 ú U00FA ù U00F9<br>④ þ                                                                                                                                                                                                                                                                    |
| <b>Ladin</b><br>– / qal                 |                         | ① –<br>② –<br>③ á U00E1 à U00E0 â U00E2 é U00E9 è U00E8<br>ê U00EA í U00ED ì U00EC î U00EE ó U00F3<br>ò U00F2 ô U00F4 ú U00FA ù U00F9 û U00FB<br>④ þ                                                                                                                                                                                                                                 |
| <b>Latin</b><br>la / lat                |                         | ① –<br>② –<br>③ –<br>④ j w þ (k)                                                                                                                                                                                                                                                                                                                                                     |
| <b>Latvian;<br/>Lettish</b><br>lv / lav |                         | ① c + č g + ģ k + ķ l + ļ n + ņ r + ŗ s + š z + ž<br>② –<br>③ ā U0101 č U010D ē U0113 ģ U???? G WITH COMMA ABOVE<br>ī U012B ķ U???? K WITH COMMA BELOW<br>ļ U???? L WITH COMMA BELOW ņ U???? N WITH COMMA BELOW<br>ō U014D ŗ U???? R WITH COMMA BELOW š U0161<br>ū U016B ž U017E<br>ģ (lowercase) corresponds to Ģ (uppercase)<br>(G WITH COMMA ABOVE/BELOW).<br>④ þ (q) (w) (x) (y) |
| <b>Lithuanian</b><br>lt / lit           |                         | ① c + č s + š z + ž y = i<br>② Ogonek before dot above.<br>③ a U0105 č U010D e U0119 é U0117 i U012F š U0161<br>ą U0173 ū U016B ž U017E<br>④ þ (q) (w) (x)                                                                                                                                                                                                                           |
| <b>Luxembur-<br/>gish</b><br>lb / ltz   |                         | ① –<br>② –<br>③ ä U00E4 é U00E9 ë U00EB #1 U???? M WITH CARON<br>#2 U???? N WITH CARON ö U00F6 ù U00FC<br>④ þ                                                                                                                                                                                                                                                                        |
| <b>Maltese</b><br>mt / mlt              |                         | ① f + ġ h + ħ n + għ y + ż ċ = c<br>② –<br>③ ċ U010B ġ U0121 ħ U0127 ż U017C (à) U00E0<br>(è) U00E8 (i) U00EC (i) U00EE (ò) U00F2 (ù) U00F9<br>④ þ (c) (y)                                                                                                                                                                                                                           |
| <b>Ndebele</b><br>– / qan (nde,<br>nbl) |                         | ① –<br>② –<br>③ –<br>④ h þ                                                                                                                                                                                                                                                                                                                                                           |

Table F.1 (continued)

| Language                                  | Variants;<br>References           | Deviations and other information on alphabet<br>(① first level; ② other levels; ③ native letters; ④ non-native)                                                                                                                                                                             |
|-------------------------------------------|-----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Norwegian</b><br>no / nor              | Bokmål and<br>Nynorsk;<br>NS 4103 | ① z + æ + ø + å    ü = y    ŷ = y    ä = æ    ö = ø    ő = ø<br>þ → th<br>Digits after letters.<br>② –<br>③ à U00E0    å U00E5    æ U00E6    é U00E9    è U00E8<br>ê U00EA    ô U00F3    ò U00F2    ô U00F4    ø U00F8<br>(ä) U00E4    (ö) U00F6    (ü) U00FC<br>④ þ                        |
| <b>Occitan;<br/>Provençal</b><br>oc / oci |                                   | ① –<br>② –<br>③ ç U00E7    ï U00EF<br>④ k    w    þ                                                                                                                                                                                                                                         |
| <b>Polish</b><br>pl / pol                 |                                   | ① a + ą    c + ć    e + ę    l + ł    n + ŋ    o + ó    s + ś    z + ź + ż<br>② –<br>③ ą U0105    ć U0107    ę U0119    ł U0142    ŋ U0144    ó U00F3<br>ś U015B    ź U017A    ż U017C<br>④ þ    (q)    (v)    (x)                                                                          |
| <b>Portuguese</b><br>pt / por             |                                   | ① –<br>② –<br>③ á U00E1    à U00E0    â U00E2    ã U00E3    ç U00E7    é U00E9<br>è U00E8    ê U00EA    í U00ED    ó U00F3    ô U00F2    ô U00F4<br>õ U00F5    ú U00FA    ü U00FC<br>④ þ                                                                                                    |
| <b>Raeto-<br/>Romance</b><br>rm / roh     |                                   | ① –<br>② –<br>③ à U00E0    â U00E2    è U00E8    ê U00EA    ò U00F2<br>ö U00F6    ü U00FC<br>④ þ                                                                                                                                                                                            |
| <b>Romanian<sup>a</sup></b><br>ro / ron   |                                   | ① a + â + â    i + î    s + ș    t + ț<br>② –<br>③ ă U0103    â U00E2    î U00EE    ș U???? S WITH COMMA BELOW<br>ț U???? T WITH COMMA BELOW<br>④ þ                                                                                                                                         |
| <b>Sami</b><br>– / smi                    | Northern<br>Sami<br>se / sme      | ① a + á    c + č    d + đ    n + ŋ    s + š    t + t̥    z + ž + (æ) + (ø) +<br>(ǣ)    (ǣ) = (æ)    (ö) = (ø)<br>② –<br>③ á U00E1    č U010D    đ U0111    ŋ U014B    š U0161    t̥ U0167<br>ž U017E    (ǣ) U00E4    (æ) U00E6    (ö) U00F6    (ø) U00F8<br>④ þ    (q)    (w)    (x)    (y) |
|                                           | Lule Sami                         | ① a + á    n + ŋ    z + ǣ + ǣ<br>② –<br>③ á U00E1    ǣ U00E5    ä U00E4    ŋ U0144<br>④ þ    (c)    (q)    (w)    (x)    (y)    (z)                                                                                                                                                         |

Table F.1 (continued)

| Language                            | Variants;<br>References | Deviations and other information on alphabet<br>(① first level; ② other levels; ③ native letters; ④ non-native)                                                                                                                                                                                                                                                                                                                                                 |
|-------------------------------------|-------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                     | Inari Sami              | ① c + č d + đ s + š z + ž + ä á = ä<br>② –<br>③ á U00E1 â U00E2 ä U00E4 č U010D đ U0111 š U0161<br>ž U017E<br>④ þ (q) (w) (x)                                                                                                                                                                                                                                                                                                                                   |
|                                     | Southern Sami           | ① z + æ + ø + å (ä) = æ (ö) = ø<br>② –<br>③ å U00E5 æ U00E6 ø U00F8 (ä) U00E4 (ö) U00F6<br>④ þ (c) (q) (w) (x) (z)                                                                                                                                                                                                                                                                                                                                              |
| <b>Scottish Gaelic</b><br>gd / gla  |                         | ① –<br>② –<br>③ á U00E1 à U00E0 é U00E9 è U00E8 í U00ED ì U00EC<br>ó U00F3 ò U00F2 ù U00F9<br>④ þ (j) (k) (q) (v) (w) (x) (y) (z)                                                                                                                                                                                                                                                                                                                               |
| <b>Serbian</b><br>sr / srp          |                         | ① c + č + ć d + dž + đ l + lj n + nj s + š z + ž<br>② –<br>③ č U010D ć U0107 đ U0111 š U0161 ž U017E<br>④ þ (q) (w) (x) (y)                                                                                                                                                                                                                                                                                                                                     |
| <b>Setswana; Tswana</b><br>tn / tsn |                         | ① e + ê o + ô s + š<br>② –<br>③ ê U00EA ô U00F4 š U0161<br>④ c q v x z þ                                                                                                                                                                                                                                                                                                                                                                                        |
| <b>Siswati; Swati</b><br>ss / ssw   |                         | ① –<br>② –<br>③ á U00E1 é U00E9 ó U00F3<br>④ q r x þ                                                                                                                                                                                                                                                                                                                                                                                                            |
| <b>Slovak</b><br>sk / slk           |                         | ① a + ä c + č h + ch r + ř s + š z + ž<br>② –<br>③ á U00E1 ä U00E4 č U010D d' U???? Ď U010E é U00E9<br>í U00ED ñ U0148 ó U00F3 ô U00F4 ř U0155 ŕ U0159<br>š U0161 ť U???? Ť U0164 ú U00FA ý U00FD ž U017E<br>d' (lowercase) corresponds to Ď (uppercase) and is sometimes represented as U010F LATIN SMALL LETTER D WITH CARON.<br>ť (lowercase) corresponds to Ť (uppercase) and is sometimes represented as U0165 LATIN SMALL LETTER T WITH CARON.<br>④ þ (w) |
| <b>Slovenian</b><br>sl / slv        |                         | ① c + č s + š z + ž<br>② –<br>③ č U010D é U00E9 ê U00EA ó U00F3 ô U00F4 š U0161<br>ž U017E<br>④ þ (q) (w) (x) (y)                                                                                                                                                                                                                                                                                                                                               |



Table F.1 (continued)

| Language                                  | Variants;<br>References                                         | Deviations and other information on alphabet<br>(① first level; ② other levels; ③ native letters; ④ non-native)                                                                                                                                                                                                   |
|-------------------------------------------|-----------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Somali</b><br>so / som                 |                                                                 | ① d + dh      k + kh      s + sh<br>② –<br>③ –<br>④ p      v      z      þ                                                                                                                                                                                                                                        |
| <b>Sotho,<br/>Northern</b><br>– / nso     |                                                                 | ① e + ê      o + ô      s + š<br>② –<br>③ ê U00EA      ô U00F4      š U0161<br>④ þ      (c)      (q)      (v)      (z)                                                                                                                                                                                            |
| <b>Spanish;<br/>Castilian</b><br>es / spa | Real Academia de la Lengua Española (new alphabetical sequence) | ① n + ñ<br>② –<br>③ á U00E1      é U00E9      í U00ED      ñ U00F1      ó U00F3      ú U00FA<br>ü U00FC<br>④ þ      (k)      (w)                                                                                                                                                                                  |
|                                           | Traditional alphabetical sequence                               | ① c + ch      l + ll      n + ñ<br>② –<br>③ á U00E1      é U00E9      í U00ED      ñ U00F1      ó U00F3      ú U00FA<br>ü U00FC<br>④ þ      (k)      (w)                                                                                                                                                          |
| <b>Swahili;<br/>Kiswahili</b><br>sw / swa |                                                                 | ① b + ch + d + dh      g + gh      n + ng' + ny      s + sh + t + th<br>② –<br>③ –<br>④ þ                                                                                                                                                                                                                         |
| <b>Swedish</b><br>sv / swe                | TN 34                                                           | ① z + å + ä + ö      v = w      æ = a      ü = y      ø = ö<br>② –<br>③ ä U00E4      å U00E5      é U00E9      ö U00F6      (æ) U00E6<br>(ø) U00F8      (ü) U00FC<br>④ þ                                                                                                                                          |
| <b>Tsonga</b><br>ts / tso                 |                                                                 | ① n + n'<br>② –<br>③ n' U???? N FOLLOWED BY APOSTROPHE<br>④ þ                                                                                                                                                                                                                                                     |
| <b>Turkish</b><br>tr / tur                |                                                                 | ① c + ç      g + ğ      h + ı + i      o + ö      s + ş      u + ü<br>② –<br>③ â U00E2      ç U00E7      ğ U011F      ı U0131      İ U0130      ö U00F6<br>ş U015F      û U00FB      ü U00FC<br>For details about ı/I and i/İ, see 5.2, item b), second paragraph after NOTE 3.<br>④ þ      (q)      (w)      (x) |

Table F.1 (continued)

| Language                      | Variants;<br>References | Deviations and other information on alphabet<br>(① first level; ② other levels; ③ native letters; ④ non-native)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|-------------------------------|-------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Venda</b><br>ve / ven      |                         | ① b + #1 + d      k + #2 + l      m + #3 + n + #4      s + #5 + t<br>② –<br>③ #1 U1E13 D WITH CIRCUMFLEX BELOW<br>#2 U1E3D L WITH CIRCUMFLEX BELOW<br>#3 U1E4B N WITH CIRCUMFLEX BELOW<br>#4 U1E45 N WITH DOT ABOVE      #5 U1E71 T WITH CIRCUMFLEX BELOW<br>④ c      j      q      p                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
| <b>Vietnamese</b><br>vi / vie |                         | ① a + ã + â      c + ch      d + đ      e + ê      g + gh + gi      k + kh<br>n + ng + ngh + nh      o + ô + #31      t + th + tr      u + #40<br>② Ordering according to diacritical marks indicating tone: grave, acute, hook above, tilde, dot below.<br>③ à U00E0      á U00E1      #01 U1EA3 A WITH HOOK ABOVE      ã U00E3<br>#02 U1EA1 A WITH DOT BELOW      ă U0103      #03 U1EB1 A WITH<br>BREVE AND GRAVE      #04 U1EAF A WITH BREVE AND ACUTE      #0<br>5 U1EB3 A WITH BREVE AND HOOK ABOVE      #06 U1EB5 A WITH BRE<br>VE AND TILDE      #07 U1EB7 A WITH BREVE AND TILDE      â U00E2<br>#08 U1EA7 A WITH CIRCUMFLEX AND GRAVE<br>#09 U1EA5 A WITH CIRCUMFLEX AND ACUTE      #10 U1EA9 A WITH CIR-<br>CUMFLEX AND HOOK ABOVE      #11 U1EAB A WITH CIRCUMFLEX AND TILDE<br>#12 U1EAD A WITH CIRCUMFLEX AND DOT BELOW      đ U0111      è U00E8<br>é U00E9      #13 U1EBB E WITH HOOK ABOVE      #14 U1EBD E WITH TILDE<br>#15 U1EB9 E WITH DOT BELOW      ê U00EA<br>#16 U1EC1 E WITH CIRCUMFLEX AND GRAVE      #17 U1EBF E WITH CIR-<br>CUMFLEX AND ACUTE      #18 U1EC3 E WITH CIRCUMFLEX AND HOOK ABOVE<br>#19 U1EC5 E WITH CIRCUMFLEX AND TILDE      #20 U1EC7 E WITH CIR-<br>CUMFLEX AND DOT BELOW      ì U00EC      í U00ED      #21 U1EC9 I<br>WITH HOOK ABOVE      #22 U0129 I WITH TILDE      #23 U1ECB I WIT<br>H DOT BELOW      ò U00F2      ó U00F3      #24 U1ECF O WITH HOOK<br>ABOVE      ô U00F5      #25 U1ECD O WITH DOT BELOW      ô U00F4<br>#26 U1ED3 O WITH CIRCUMFLEX AND GRAVE      #27 U1ED1 O WITH CIR-<br>CUMFLEX AND ACUTE      #28 U1ED5 O WITH CIRCUMFLEX AND HOOK ABOVE<br>#29 U1ED7 O WITH CIRCUMFLEX AND TILDE<br>#30 U1ED9 O WITH CIRCUMFLEX AND DOT BELOW<br>#31 U01A1 O WITH HORN      #32 U1EDD O WITH HORN AND GRAVE      #3<br>3 U1EDB O WITH HORN AND ACUTE      #34 U1EDF O WITH HORN AND HO<br>OK ABOVE      #35 U1EE1 O WITH HORN AND TILDE      #36 U1EE3 O WIT<br>H HORN AND DOT BELOW      ù U00F9      ú U00FA      #37 U1EE7 U WI<br>TH HOOK ABOVE      #38 U0169 U WITH TILDE      #39 U1EE5 U WITH DO<br>T BELOW      #40 U01B0 U WITH HORN      #41 U1EEB U WITH HORN AN<br>D GRAVE      #42 U1EE9 U WITH HORN AND ACUTE      #43 U1EED U WIT<br>H HORN AND HOOK ABOVE      #44 U1EEF U WITH HORN AND TILDE      #4<br>5 U1EF1 U WITH HORN AND DOT BELOW      #46 U1EF3 Y WITH GRAVE<br>ý U00FD      #47 U1EF7 Y WITH HOOK ABOVE      #48 U1EF9 Y WITH TILDE<br>#49 U1EF5 Y WITH DOT BELOW<br>④ f      j      v      z      p |

Table F.1 (continued)

| Language                 | Variants;<br>References | Deviations and other information on alphabet<br>(① first level; ② other levels; ③ native letters; ④ non-native)                                                                                                                                                                                                                                                                                                                                                                                                        |
|--------------------------|-------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Welsh</b><br>cy / cym |                         | ① c + ch    d + dd    f + ff    g + ng    l + ll    p + ph    r + rh<br>t + th<br>② –<br>③ á U00E1    à U00E0    â U00E2    ä U00E4    é U00E9    è U00E8<br>ê U00EA    ë U00EB    í U00ED    ì U00EC    î U00EE    ï U00EF<br>ó U00F3    ô U00F2    ô U00F4    ö U00F6    ú U00FA<br>ù U00F9    û U00FB    ü U00FC    #1 U1E83 w WITH ACUTE<br>#2 U1E81 w WITH GRAVE    ŵ U0175    #3 U1E85 w WITH DIAERESIS<br>ý U00FD    #4 U1EF3 Y WITH GRAVE    #5 U0177 Y WITH CIRCUMFLEX<br>ÿ U00FF<br>④ þ    (q)    (v)    (x) |
| <b>Xhosa</b><br>xh / xho |                         | ① –<br>② –<br>③ –<br>④ þ                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| <b>Zulu</b><br>zu / zul  |                         | ① –<br>② –<br>③ –<br>④ þ    (r)                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |

Table F.2 — Character repertoires of some transliteration standards

| Language,<br>Script                  | References | Character repertoire                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|--------------------------------------|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Greek</b><br>[el]                 | ISO 843    | Table 1: a    v    g    d    e    z    ī U012B (or i-)    th<br>i    k    l    m    n    x    o    p    r    s    t    y<br>f    ch    ps    ō U014D (or o-)<br>Table 2: a    ai    áí    aĩ    av    af    ay    v    g    ng<br>gk    nx    nch    d    e    ei    éi    eĩ    ev    ef    ey<br>z    i    iv    if    iy    th    i    k    l    m    b    mp<br>n    nt    x    o    oi    ói    oĩ    ou    oy    p    r<br>s    t    y    yi    f    ch    ps    o                                                                                                                                                                       |
| <b>Cyrillic<sup>a</sup></b><br>[Cyr] | ISO 9      | Table 1: a    b    v    g    d    đ U0111<br>#1 U01F5 G WITH ACUTE    e    ë U00EB    ê U00EA    ž U017E<br>z    #2 U1E91 Z WITH CIRCUMFLEX    i    ì U00EC    ĭ U00EF<br>#3 U01F0 J WITH CARON    j    k    l<br>#4 U???? L WITH CIRCUMFLEX    m    n<br>#5 U???? N WITH CIRCUMFLEX    o    p    r    s    t<br>ć U0107    #6 U1E31 K WITH ACUTE    u    ŭ U016D    f<br>h    c    č U010D    #7 U???? D WITH CIRCUMFLEX    š U0161<br>š U015D    " U0022    y    ' U0027    è U00E8    û U00FB<br>â U00E2    ' U02BC<br>Table 2: #1 U???? G WITH GRAVE    ě U011B    #2 U01CE A WITH CARON<br>#3 U???? F WITH GRAVE    #4 U1EF3 Y WITH GRAVE |

<sup>a</sup> The symbol *Cyr* is used for the Cyrillic alphabet.

Table F.2 (continued)

| Language, Script                                               | References | Character repertoire                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|----------------------------------------------------------------|------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                                                                |            | <p><i>Table 3:</i>    a    ä U00E4    #01 U???? A WITH DIAERESIS AND DOT BELOW<br/> ä U0103    ā U0101    æ U00E6    á U00E1    å U00E5    b    v<br/> g    #02 U01F5 G WITH ACUTE    ġ U0121    ĝ U011F    #03 U1E2<br/> 5 H WITH DOT BELOW    d    e    ě U0115    ě U00EB    ž U017E<br/> #04 U???? Z WITH CARON AND CEDILLA<br/> #05 U???? Z WITH MACRON    #06 U???? Z WITH BREVE    z<br/> #07 U???? Z WITH CEDILLA    #08 U???? Z WITH DIAERESIS<br/> ž U017A    i    ī U012B    í U00ED    î U00EE    j    ï U00EC<br/> ï U00EF    #09 U01D0 I WITH CARON    #10 U01F0 J WITH CARON<br/> #11 U???? J WITH ACUTE    k    #12 U???? K WITH DOT BELOW<br/> w    #13 U???? K WITH CIRCUMFLEX    #14 U???? K WITH CARON<br/> #15 U???? K WITH MACRON    #16 U???? K WITH COMMA BELOW<br/> w    #17 U???? K WITH GRAVE    q    l    #18 U013A L WITH ACUTE<br/> #19 U???? L WITH COMMA BELOW    m    n    #20 U???? N WI<br/> TH COMMA BELOW    #21 U1E47 N WITH DOT BELOW    #22 U1<br/> E45 N WITH DOT ABOVE    #23 U???? N WITH GRAVE    ñ U0144<br/> #24 U???? N WITH BREVE    #25 U???? N WITH MACRON    o<br/> ö U00F6    ô U00F4    õ U0151    #26 U???? O WITH DIAERE-<br/> SIS AND DOT BELOW    ò U00F2    ó U00F3    ô U014D    œ U0153<br/> p    #27 U1E55 P WITH ACUTE    #28 U???? P WITH GRAVE    r<br/> s    #29 U???? S WITH COMMA BELOW    #30 U???? S WITH GRAVE<br/> t    #31 U???? T WITH GRAVE    #32 U0165 T WITH CARON    #3<br/> 3 U???? T WITH COMMA BELOW    u    ü U00FC    ŭ U016B<br/> ű U016D    ů U0171    ú U00FA    #34 U???? U WITH DIAERE-<br/> SIS AND DOT BELOW    ù U00F9    #35 U???? U WITH DOT ABOVE<br/> E    #36 U???? U WITH MACRON AND DOT BELOW    w    f    h<br/> #37 U1E29 H WITH CEDILLA    c    #38 U???? X WITH MACRON<br/> #39 U???? D WITH CIRCUMFLEX    č U010D    ç U00E7    #40 U?<br/> ??? C WITH DOT BELOW    #41 U???? C WITH DIAERESIS    ĉ U0109<br/> #42 U???? C WITH GRAVE    #43 U???? C WITH BREVE    #44 U???? C<br/> WITH BREVE AND COMMA BELOW    š U0161    ŝ U015D    “ U0022<br/> y    ÿ U00FF    #45 U???? Y WITH MACRON    ' U0027    è U00E8<br/> #46 U???? A WITH DOUBLE ACUTE    à U00E0    â U00FB    ã U016F<br/> â U00E2</p> |
| Arabic<br>[ar]                                                 | ISO 233    | <p><i>Table 1:</i>    ' U02BC    'ā U02BC+00E2    b    t<br/> #01 U1E6F T WITH LINE BELOW    #02 U01E7 G WITH CARON    #03 U1<br/> E25 H WITH DOT BELOW    #04 U1E96 H WITH LINE BELOW    d    #0<br/> 5 U1E0F D WITH LINE BELOW    r    z    s    š U0161    #0<br/> 6 U1E63 S WITH DOT BELOW    #07 U1E0D D WITH DOT BELOW<br/> w    #08 U1E6D T WITH DOT BELOW    #09 U1E93 Z WITH DOT BELOW<br/> w    ' U02BB    ġ U0121    f    q    k    l    m    n    h<br/> #10 U1E97 T WITH DIAERESIS    w    y    #11 U1EF3 Y WITH GRAVE</p> <p><i>Table 2:</i>    a    a'    á' U00E1+02BC    á U00E1    ā U0101    u<br/> ú U00FA    i    í U00ED</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |
|                                                                | ISO 233-2  | <p><i>Table 1:</i>    ' U02BC    b    t    #01 U1E6F T WITH LINE BELOW<br/> #02 U01E7 G WITH CARON    #03 U1E25 H WITH DOT BELOW    #04 U1<br/> E96 H WITH LINE BELOW    d    #05 U1E0F D WITH LINE BELOW    r<br/> z    s    š U0161    #06 U1E63 S WITH DOT BELOW    #07 U1E0D D<br/> WITH DOT BELOW    #08 U1E6D T WITH DOT BELOW    #09 U1E93 Z W<br/> ITH DOT BELOW    ' U02BB    ġ U0121    f    q    k    l    m<br/> h    #10 U1E97 T WITH DIAERESIS    w    y</p> <p><i>Table 2:</i>    a    ā U0101    â U00E2    á U00E1    u    ŭ U016B<br/> i<br/> ī U012B</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <sup>a</sup> The symbol Cyr is used for the Cyrillic alphabet. |            |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |

Table F.2 (continued)

| Language,<br>Script                                            | References | Character repertoire                                                                                                                                                                                                                                                                                                                                                                                |
|----------------------------------------------------------------|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Hebrew</b><br>[he]                                          | ISO 259-2  | <p><i>Table 1:</i> ' U02BC b #1 U1E03 B WITH DOT ABOVE g d h<br/> #2 U1E23 H WITH DOT ABOVE w z #3 U1E25 H WITH DO<br/> T BELOW #4 U1E6D T WITH DOT BELOW y #5 U???? K WI<br/> TH DOT ABOVE l m n s 'U01BB p #6 U1<br/> E57 P WITH DOT ABOVE #7 U1E63 S WITH DOT BELOW q r<br/> #8 U???? S WITH GRAVE š U0161 š U015B t</p> <p><i>Table 2:</i> a e i o u ŵ U0175<br/> #1 U1E87 W WITH DOT ABOVE</p> |
| <sup>a</sup> The symbol Cyr is used for the Cyrillic alphabet. |            |                                                                                                                                                                                                                                                                                                                                                                                                     |

## Annex G (informative)

### Formal description of the rules of the main body of this document

```
%% ISO 12199 Table
section ISO12199CollatingSymbols
%% This table is based on ISO14651_1999_TABLE1
%%
%% Digits
<0>
<1>
<2>
<3>
<4>
<5>
<6>
<7>
<8>
<9>
%% Latin letters
<a>
<b>
<c>
<d>
<e>
<f>
<g>
<h>
<i>
<j>
<k>
<l>
<m>
<n>
<o>
<p>
<q>
<r>
<s>
<t>
<u>
<v>
<w>
<x>
<y>
<z>
<thorn>
%% Collating symbols for the second ordering level key
<BLANK>
<ACUTE>
<GRAVE>
<BREVE>
<CIRCUMFLEX>
<CIRCUMFLEX_BELOW>
<CARON>
<RING_ABOVE>
<DIAERESIS>
<DOUBLE_ACUTE>
<HOOK_ABOVE>
<TILDE>
<DOT_ABOVE>
<DOT_BELOW>
<CEDILLA>
<COMMA>
```

<OGONEK>  
 <MACRON>  
 <MACRON\_BELOW>  
 <HORN>  
 <MODIFIED1>  
 <MODIFIED2>  
 <MODIFIED3>  
 <SUBSTITUTE>  
 %% Collating symbols for the third ordering level key  
 <LOWERCASE>  
 <UPPERCASE>  
 %%  
 section ISO12199Digits  
 <U0030> <0>; <BLANK>; <LOWERCASE>; <U0030> % DIGIT ZERO  
 <U2070> <0>; <BLANK>; <LOWERCASE>; <U2070> % SUPERSCRIPT ZERO  
 <U2080> <0>; <BLANK>; <LOWERCASE>; <U2080> % SUBSCRIPT ZERO  
 <U0031> <1>; <BLANK>; <LOWERCASE>; <U0031> % DIGIT ONE  
 <U00B9> <1>; <BLANK>; <LOWERCASE>; <U00B9> % SUPERSCRIPT ONE  
 <U2081> <1>; <BLANK>; <LOWERCASE>; <U2081> % SUBSCRIPT ONE  
 <U0032> <2>; <BLANK>; <LOWERCASE>; <U0032> % DIGIT TWO  
 <U00B2> <2>; <BLANK>; <LOWERCASE>; <U00B2> % SUPERSCRIPT TWO  
 <U2082> <2>; <BLANK>; <LOWERCASE>; <U2082> % SUBSCRIPT TWO  
 <U0033> <3>; <BLANK>; <LOWERCASE>; <U0033> % DIGIT THREE  
 <U00B3> <3>; <BLANK>; <LOWERCASE>; <U00B3> % SUPERSCRIPT THREE  
 <U2083> <3>; <BLANK>; <LOWERCASE>; <U2083> % SUBSCRIPT THREE  
 <U0034> <4>; <BLANK>; <LOWERCASE>; <U0034> % DIGIT FOUR  
 <U2074> <4>; <BLANK>; <LOWERCASE>; <U2074> % SUPERSCRIPT FOUR  
 <U2084> <4>; <BLANK>; <LOWERCASE>; <U2084> % SUBSCRIPT FOUR  
 <U0035> <5>; <BLANK>; <LOWERCASE>; <U0035> % DIGIT FIVE  
 <U2075> <5>; <BLANK>; <LOWERCASE>; <U2075> % SUPERSCRIPT FIVE  
 <U2085> <5>; <BLANK>; <LOWERCASE>; <U2085> % SUBSCRIPT FIVE  
 <U0036> <6>; <BLANK>; <LOWERCASE>; <U0036> % DIGIT SIX  
 <U2076> <6>; <BLANK>; <LOWERCASE>; <U2076> % SUPERSCRIPT SIX  
 <U2086> <6>; <BLANK>; <LOWERCASE>; <U2086> % SUBSCRIPT SIX  
 <U0037> <7>; <BLANK>; <LOWERCASE>; <U0037> % DIGIT SEVEN  
 <U2077> <7>; <BLANK>; <LOWERCASE>; <U2077> % SUPERSCRIPT SEVEN  
 <U2087> <7>; <BLANK>; <LOWERCASE>; <U2087> % SUBSCRIPT SEVEN  
 <U0038> <8>; <BLANK>; <LOWERCASE>; <U0038> % DIGIT EIGHT  
 <U2078> <8>; <BLANK>; <LOWERCASE>; <U2078> % SUPERSCRIPT EIGHT  
 <U2088> <8>; <BLANK>; <LOWERCASE>; <U2088> % SUBSCRIPT EIGHT  
 <U0039> <9>; <BLANK>; <LOWERCASE>; <U0039> % DIGIT NINE  
 <U2079> <9>; <BLANK>; <LOWERCASE>; <U2079> % SUPERSCRIPT NINE  
 <U2089> <9>; <BLANK>; <LOWERCASE>; <U2089> % SUBSCRIPT NINE  
 %%  
 section ISO12199LatinLetters  
 <U0061> <a>; <BLANK>; <LOWERCASE>; <U0061> % LATIN SMALL LETTER A  
 <U0041> <a>; <BLANK>; <UPPERCASE>; <U0041> % LATIN CAPITAL LETTER A  
 <U00E1> <a>; <ACUTE>; <LOWERCASE>; <U00E1> % LATIN SMALL LETTER A WITH ACUTE  
 <U00C1> <a>; <ACUTE>; <UPPERCASE>; <U00C1> % LATIN CAPITAL LETTER A WITH ACUTE  
 <U00E0> <a>; <GRAVE>; <LOWERCASE>; <U00E0> % LATIN SMALL LETTER A WITH GRAVE  
 <U00C0> <a>; <GRAVE>; <UPPERCASE>; <U00C0> % LATIN CAPITAL LETTER A WITH GRAVE  
 <U0103> <a>; <BREVE>; <LOWERCASE>; <U0103> % LATIN SMALL LETTER A WITH BREVE  
 <U0102> <a>; <BREVE>; <UPPERCASE>; <U0102> % LATIN CAPITAL LETTER A WITH BREVE  
 <U1EAF> <a>; <BREVE><ACUTE>; <LOWERCASE>; <U1EAF> % LATIN SMALL LETTER A WITH BREVE AND ACUTE  
 <U1EAE> <a>; <BREVE><ACUTE>; <UPPERCASE>; <U1EAE> % LATIN CAPITAL LETTER A WITH BREVE AND ACUTE  
 <U1EB1> <a>; <BREVE><GRAVE>; <LOWERCASE>; <U1EB1> % LATIN SMALL LETTER A WITH BREVE AND GRAVE  
 <U1EB0> <a>; <BREVE><GRAVE>; <UPPERCASE>; <U1EB0> % LATIN CAPITAL LETTER A WITH BREVE AND GRAVE  
 <U1EB3> <a>; <BREVE><HOOK\_ABOVE>; <LOWERCASE>; <U1EB3> % LATIN SMALL LETTER A WITH BREVE AND HOOK ABOVE  
 <U1EB2> <a>; <BREVE><HOOK\_ABOVE>; <UPPERCASE>; <U1EB2> % LATIN CAPITAL LETTER A WITH BREVE AND HOOK ABOVE  
 <U1EB5> <a>; <BREVE><TILDE>; <LOWERCASE>; <U1EB5> % LATIN SMALL LETTER A WITH BREVE AND TILDE  
 <U1EB4> <a>; <BREVE><TILDE>; <UPPERCASE>; <U1EB4> % LATIN CAPITAL LETTER A WITH BREVE AND TILDE  
 <U1EB7> <a>; <BREVE><DOT\_BELOW>; <LOWERCASE>; <U1EB7> % LATIN SMALL LETTER A WITH BREVE AND DOT BELOW

<U1EB6> <a>; <BREVE><DOT\_BELOW>; <UPPERCASE>; <U1EB6> % LATIN CAPITAL LETTER A WITH BREVE AND DOT BELOW  
 <U00E2> <a>; <CIRCUMFLEX>; <LOWERCASE>; <U00E2> % LATIN SMALL LETTER A WITH CIRCUMFLEX  
 <U00C2> <a>; <CIRCUMFLEX>; <UPPERCASE>; <U00C2> % LATIN CAPITAL LETTER A WITH CIRCUMFLEX  
 <U1EA5> <a>; <CIRCUMFLEX><ACUTE>; <LOWERCASE>; <U1EA5> % LATIN SMALL LETTER A WITH CIRCUMFLEX AND ACUTE  
 <U1EA4> <a>; <CIRCUMFLEX><ACUTE>; <UPPERCASE>; <U1EA4> % LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND ACUTE  
 <U1EA7> <a>; <CIRCUMFLEX><GRAVE>; <LOWERCASE>; <U1EA7> % LATIN SMALL LETTER A WITH CIRCUMFLEX AND GRAVE  
 <U1EA6> <a>; <CIRCUMFLEX><GRAVE>; <UPPERCASE>; <U1EA6> % LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND GRAVE  
 <U1EA9> <a>; <CIRCUMFLEX><HOOK\_ABOVE>; <LOWERCASE>; <U1EA9> % LATIN SMALL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE  
 <U1EA8> <a>; <CIRCUMFLEX><HOOK\_ABOVE>; <UPPERCASE>; <U1EA8> % LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND HOOK ABOVE  
 <U1EAB> <a>; <CIRCUMFLEX><TILDE>; <LOWERCASE>; <U1EAB> % LATIN SMALL LETTER A WITH CIRCUMFLEX AND TILDE  
 <U1EAA> <a>; <CIRCUMFLEX><TILDE>; <UPPERCASE>; <U1EAA> % LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND TILDE  
 <U1EAD> <a>; <CIRCUMFLEX><DOT\_BELOW>; <LOWERCASE>; <U1EAD> % LATIN SMALL LETTER A WITH CIRCUMFLEX AND DOT BELOW  
 <U1EAC> <a>; <CIRCUMFLEX><DOT\_BELOW>; <UPPERCASE>; <U1EAC> % LATIN CAPITAL LETTER A WITH CIRCUMFLEX AND DOT BELOW  
 <U01CE> <a>; <CARON>; <LOWERCASE>; <U01CE> % LATIN SMALL LETTER A WITH CARON  
 <U01CD> <a>; <CARON>; <UPPERCASE>; <U01CD> % LATIN CAPITAL LETTER A WITH CARON  
 <U00E5> <a>; <RING\_ABOVE>; <LOWERCASE>; <U00E5> % LATIN SMALL LETTER A WITH RING ABOVE  
 <U00C5> <a>; <RING\_ABOVE>; <UPPERCASE>; <U00C5> % LATIN CAPITAL LETTER A WITH RING ABOVE  
 <U01FB> <a>; <RING\_ABOVE><ACUTE>; <LOWERCASE>; <U01FB> % LATIN SMALL LETTER A WITH RING ABOVE AND ACUTE  
 <U01FA> <a>; <RING\_ABOVE><ACUTE>; <UPPERCASE>; <U01FA> % LATIN CAPITAL LETTER A WITH RING ABOVE AND ACUTE  
 <U00E4> <a>; <DIAERESIS>; <LOWERCASE>; <U00E4> % LATIN SMALL LETTER A WITH DIAERESIS  
 <U00C4> <a>; <DIAERESIS>; <UPPERCASE>; <U00C4> % LATIN CAPITAL LETTER A WITH DIAERESIS  
 <U01DF> <a>; <DIAERESIS><MACRON>; <LOWERCASE>; <U01DF> % LATIN SMALL LETTER A WITH DIAERESIS AND MACRON  
 <U01DE> <a>; <DIAERESIS><MACRON>; <UPPERCASE>; <U01DE> % LATIN CAPITAL LETTER A WITH DIAERESIS AND MACRON  
 <U1EA3> <a>; <HOOK\_ABOVE>; <LOWERCASE>; <U1EA3> % LATIN SMALL LETTER A WITH HOOK ABOVE  
 <U1EA2> <a>; <HOOK\_ABOVE>; <UPPERCASE>; <U1EA2> % LATIN CAPITAL LETTER A WITH HOOK ABOVE  
 <U00E3> <a>; <TILDE>; <LOWERCASE>; <U00E3> % LATIN SMALL LETTER A WITH TILDE  
 <U00C3> <a>; <TILDE>; <UPPERCASE>; <U00C3> % LATIN CAPITAL LETTER A WITH TILDE  
 <U1EA1> <a>; <DOT\_BELOW>; <LOWERCASE>; <U1EA1> % LATIN SMALL LETTER A WITH DOT BELOW  
 <U1EA0> <a>; <DOT\_BELOW>; <UPPERCASE>; <U1EA0> % LATIN CAPITAL LETTER A WITH DOT BELOW  
 <U0105> <a>; <OGONEK>; <LOWERCASE>; <U0105> % LATIN SMALL LETTER A WITH OGONEK  
 <U0104> <a>; <OGONEK>; <UPPERCASE>; <U0104> % LATIN CAPITAL LETTER A WITH OGONEK  
 <U0101> <a>; <MACRON>; <LOWERCASE>; <U0101> % LATIN SMALL LETTER A WITH MACRON  
 <U0100> <a>; <MACRON>; <UPPERCASE>; <U0100> % LATIN CAPITAL LETTER A WITH MACRON  
 <U00E6> "<a><e>"; <MODIFIED1><BLANK>; <LOWERCASE><LOWERCASE>; <U0061><U0065> % LATIN SMALL LETTER AE  
 <U00C6> "<a><e>"; <MODIFIED1><BLANK>; <UPPERCASE><UPPERCASE>; <U0041><U0045> % LATIN CAPITAL LETTER AE  
 <U01FD> "<a><e>"; <MODIFIED1><ACUTE>; <LOWERCASE><LOWERCASE>; <U0061><U0065> % LATIN SMALL LETTER AE WITH ACUTE  
 <U01FC> "<a><e>"; <MODIFIED1><ACUTE>; <UPPERCASE><UPPERCASE>; <U0041><U0045> % LATIN CAPITAL LETTER AE WITH ACUTE  
 <U01E3> "<a><e>"; <MODIFIED1><MACRON>; <LOWERCASE><LOWERCASE>; <U0061><U0065> % LATIN SMALL LETTER AE WITH MACRON  
 <U01E2> "<a><e>"; <MODIFIED1><MACRON>; <UPPERCASE><UPPERCASE>; <U0041><U0045> % LATIN CAPITAL LETTER AE WITH MACRON  
 <U0062> <b>; <BLANK>; <LOWERCASE>; <U0062> % LATIN SMALL LETTER B  
 <U0042> <b>; <BLANK>; <UPPERCASE>; <U0042> % LATIN CAPITAL LETTER B  
 <U1E03> <b>; <DOT\_ABOVE>; <LOWERCASE>; <U1E03> % LATIN SMALL LETTER B WITH DOT ABOVE  
 <U1E02> <b>; <DOT\_ABOVE>; <UPPERCASE>; <U1E02> % LATIN CAPITAL LETTER B WITH DOT ABOVE  
 <U1E05> <b>; <DOT\_BELOW>; <LOWERCASE>; <U1E05> % LATIN SMALL LETTER B WITH DOT BELOW  
 <U1E04> <b>; <DOT\_BELOW>; <UPPERCASE>; <U1E04> % LATIN CAPITAL LETTER B WITH DOT BELOW  
 <U0253> <b>; <MODIFIED1>; <LOWERCASE>; <U0253> % LATIN SMALL LETTER B WITH HOOK  
 <U0181> <b>; <MODIFIED1>; <UPPERCASE>; <U0181> % LATIN CAPITAL LETTER B WITH HOOK  
 <U0063> <c>; <BLANK>; <LOWERCASE>; <U0063> % LATIN SMALL LETTER C  
 <U0043> <c>; <BLANK>; <UPPERCASE>; <U0043> % LATIN CAPITAL LETTER C  
 <U0107> <c>; <ACUTE>; <LOWERCASE>; <U0107> % LATIN SMALL LETTER C WITH ACUTE