INTERNATIONAL STANDARD

ISO/IEC 15897

First edition
1999-10-15

Information technology — Procedures for registration of cultural elements

Technologies de l'information — Procédures pour l'enregistrement des éléments culturels

Reference number
ISO/IEC 15897:1999(E)

Contents Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 3.

In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

International Standard ISO/IEC 15897 was prepared by the Technical Committee CEN/TC 304 "Character Set Technology" (as CEN ENV 12005), and it was adopted under a special "fast-track" procedure by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, in parallel with its approval by national bodies of ISO and IEC.

Annexes A, C and E form a normative part of this International Standard. Annexes B, D, F, G and H are for information only.

Introduction

Cultural differences throughout the world make it necessary to adopt IT-equipment to each local culture. Standard methods, being developed by ISO/IEC JTC1/SC22, make such adoption easier. Registering the local conventions in a globally available registry make it still easier. CEN/TC304 was the first committee to start work on such a registry and the European prestandard ENV 12005:1996 was the result. This International Standard is the result of the fast-track of DS/ENV 12005, a Danish Standards Association standards publication equivalent to CEN ENV 12005.

This International Standard sets out the procedures for registering cultural elements, both as narrative text and in a more formal manner, using the techniques of ISO/IEC 9945-2 POSIX Shell and Utilities. The registration will be free-of-charge and the results will also be freely available on the network at the address http://www.dkuug.dk/cultreg/. This will make information on cultural conventions freely and easily available to producers in the IT market. Some of these conventions can even be implemented automatically by downloading the formal POSIX specifications.

Information technology - Procedures for registration of cultural elements

1 Scope

This International Standard specifies the procedures to be followed in preparing, publishing and maintaining a register of cultural specifications for computer use, including freeform narrative cultural elements specifications, and POSIX Locales and Charmaps conforming to ISO/IEC 9945-2. The registry is in printed and electronic form, and the text of the cultural specifications are recorded in a way that is independent of any coded character set.

Each cultural element registration shall have unique identifiers in a particular standard format defined below in 6.8. A numeric identifier and a token identifier shall be assigned to each registered cultural element specification, POSIX Locale, Charmap and Repertoire-map. These identifiers are for unique identification of the cultural specification, and intended to be used with POSIX locale handling mechanisms and possibly other locale and charmap usage, such as in programming languages, database handling and communication protocols and for identification and specification by human operators.

The field of application of this International Standard is to provide reference for implementers, procurers, users, and standardization organizations, to determine cultural requirements in a given cultural environment. Registered items using certain POSIX formal specification methods can also be used by the POSIX Operating System and other software capable of using such specifications.

2 Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of this International Standard. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent editions of the normative documents indicated below. For undated references, the latest edition of the normative document referred to applies. Members of ISO and IEC maintain registers of currently valid International Standards.

ISO 639:1988, *Code for the representation of names of languages.*

ISO/IEC 646:1991, *Information technology - ISO 7-bit coded character set for information interchange.*

ISO/IEC 2022:1994, *Information technology - Character code structure and extension techniques.*

ISO 3166 (all parts), *Codes for the representation of names of countries and their subdivisions.*

ISO 4217:1995, *Codes for the representation of currencies and funds.*

ISO 8601:1988, *Data elements and interchange formats - Information interchange - Representation of dates and times.*

ISO/IEC 8824:1990, *Information technology - Open Systems Interconnection - Specification of Abstract Syntax Notation One (ASN.1).*

ISO/IEC 8825:1990, *Information technology - Open Systems Interconnection - Specification of Basic Encoding Rules for Abstract Syntax Notation One (ASN.1).*

ISO/IEC 9945-2:1993, *Information technology - Portable Operating System Interface (POSIX) Part 2: Shell and Utilities.*

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

3 Definitions

For the purposes of this International Standard, the following definitions apply.

3.1 Locale: the definition of the subset of the environment of a user that depends on language and cultural conventions (see 2.5 of the POSIX standard for a specification of the locale file format)

3.2 Charmap: a text file describing a coded character set (see 2.4 of the POSIX standard for a description of the Charmap file format)

3.3 Text File: a file that contains characters organized into one or more lines

3.4 Cultural Element: a data item for computer use that may vary dependent on language, territory, or other cultural circumstances

3.5 Cultural Specification: either a Narrative Cultural Specification, a related POSIX Locale or a POSIX Charmap or a Repertoiremap

3.6 Narrative Cultural Specification: a narrative description for computer use of culturally dependent information, further described in 6.2

3.7 Repertoiremap: a definition of a mapping between character names and characters for the repertoire of characters used in a Locale, further described in 6.9.

3.8 Profile (of a standard): a set of specifications of values of parameters in the standard, selections of optional items of the standard, or the recommendations concerning implementation-related matters of the standard.

4 Registration authority

The Registration Authority shall maintain a register of Cultural Specifications and their numeric and token identifiers. The contents of this register shall be available to ISO/IEC JTC 1 members and to the general public, both in printed and electronic form (for example, text files, postscript and/or fax format).

The ISO and IEC councils shall designate a Registration Authority for this International Standard. The initial registration Authority is designated to be:

The Danish UNIX-systems User Group (DKUUG)
Fruebjergvej 3
DK-2100 København Ø
Danmark
Fax: +45 3920-8948
email: culturalregister@dkuug.dk
web: http://www.dkuug.dk/cultreg/

With regard to the assignment of numeric and token identifiers to Cultural Specifications, and of subsequent additions of Cultural Specifications to the register, the responsibilities of the Registration Authority shall be

a) to receive from Sponsoring Authorities proposals of Cultural Specifications and associated token identifiers;

b) to ascertain that each proposal complies with clause 6;

c) in the case of a POSIX Locale, to ascertain that the POSIX Locale and the corresponding Narrative Cultural Specification are not in contradiction;

d) to circulate proposals to ISO/IEC JTC1/SC22 members and liaisons, with the proposed identifiers for a three month period for information and comments;

e) to forward the comments received to the Sponsoring Authority for possible integration in the final documents;

f) in the case of comments, to optionally receive from the Sponsoring Authority revised proposals addressing the comments received;

g) to assign to the Cultural Specification appropriate token identifiers based on the information given by the Sponsoring Authority, and to assign to the Cultural Specification the next available number to be used as a numeric identifier when the proposal complies with the rules, unless the Cultural Specification is identical to one already registered, in which case the new token identifiers shall be added to the existing registration;

h) to promulgate to all ISO/IEC JTC 1 members and liaisons the Cultural Specifications registered and the identifiers assigned;

i) to inform the appropriate Sponsoring Authority when a proposal does not comply to the rules.

The contents of an individual registration shall never be changed or deleted once it has been registered (except for name additions). Even the correction of editorial errors will make a new registration necessary. This is contrasted with a standard which must be reviewed and/or revised periodically. When a standard is revised that has been used as basis for a Narrative Cultural Specification, a POSIX Locale, Charmap, or Repertoiremap, these are not changed in the register. If a Sponsoring Authority desires recognition of such a revision, a new registration shall be made by following the normal procedures.

5 Sponsoring authorities

5.1 Proposals for registration of Cultural Specifications may be made only by the following bodies, which, for the purpose of this International Standard, are considered Sponsoring Authorities:

a) Any Member Body or Associated Member Body of CEN or ISO/IEC JTC1, for proposals related to the territories for which they have authority;

b) ISO/IEC JTC1/SC22 for proposals related to wider regions.

Sponsoring Authorities may submit proposals for registration of the types POSIX Charmaps and Repertoiremaps to support their other Cultural Specifications.

Proposals from other sources must be referred for consideration to the Sponsoring Authorities as noted below.

5.2 The responsibilities of the Sponsoring Authorities shall be:

a) to receive proposals concerning Cultural Specifications from their respective countries or organizations;

b) to make certain that proposals follow the rules for definitions outlined in clause 6;

c) in the case of a POSIX Locale, to make certain that Narrative Cultural Specification and the derived POSIX Locale are not in contradiction;

d) to effect such rationalization or coordination of all proposals under consideration as they may desire;

e) to forward to the Registration Authority those proposals that have their support;

f) to make known within their respective countries or organizations the result of the registration procedure as transmitted to them by the Registration Authority.

6 Rules for proposals

Four types of Cultural Specifications can be registered according to this International Standard:

1. Narrative Cultural Specification
2. POSIX Locale
3. POSIX Charmap
4. POSIX Repertoiremap

The relation between the types is the following:

1. The Narrative Cultural Specification shall specify cultural conventions in narrative English, and may give equivalent specifications in other languages. It may thus address issues which have not yet been codified by formal methods for specifications of cultural conventions. If parts of a Narrative Cultural Specification have been specified also in POSIX Locale or Charmap format, this Locale or Charmap shall be referenced in the specification.

Types 2, 3 and 4 are for POSIX specification of cultural elements defined in ISO/IEC 9945-2.

2. The POSIX Locale shall specify in formal POSIX syntax some aspects of a Narrative Cultural Specification, and shall refer to a corresponding Narrative Cultural Specification. The POSIX Locale shall refer to the Repertoiremap being used, and should also list a number of POSIX Charmaps that it can use.

3. The POSIX Charmap shall specify aspects of a Narrative Cultural Specification or a POSIX Locale that relate to coded character sets. A POSIX Charmap shall refer to the POSIX Repertoiremap being used, but need not refer to the POSIX Locales nor the Narrative Cultural Specifications using it.

4. The POSIX Repertoiremap is used as a tool to enable a POSIX Locale or a Narrative Cultural Specification to be independent of coded character sets, and to remove the requirement for POSIX Charmaps when registering a POSIX Locale. It need not refer to other Cultural Specifications.

NOTE: It is the intention to allow other formal specification methods in future revisions of this International Standard when they become standardized methods; for the time being these specifications can be registered as type 1.

The following rules apply when registering a Cultural Specification:

6.1 A proposal for registration of a Cultural Specification shall be submitted as a Text File. A Narrative Cultural Specification may alternatively be submitted on white paper of approximate size 297 * 210 mm, with margins of no less than 20 mm.

6.2 The format of a Narrative Cultural Specification shall be in ordinary text, and shall contain clauses describing the following cultural elements, which may also be described in a POSIX Locale:

1 Alphanumeric deterministic ordering
2 Classification of characters
3 Numeric formatting
4 Monetary formatting
5 Date and time conventions
6 Affirmative and negative answers

The Narrative Cultural Specification may also include other culturally dependent information, limited to the following list:

7 National or cultural Information Technology terminology
8 National or cultural profiles of standards
9 Character set considerations
10 Sorting and searching rules
11 Transformation of characters
12 Character properties
13 Use of special characters
14 Character rendition
15 Character inputting
16 Personal names rules
17 Inflection
18 Hyphenation
19 Spelling
20 Numbering, ordinals and measuring systems
21 Monetary amounts
22 Date and time
23 Coding of national entities
24 Telephone numbers
25 Mail addresses
26 Identification of persons and organizations
27 Electronic mail addresses
28 Payment account numbers
29 Keyboard layout
30 Man-machine dialogue
31 Paper formats
32 Typographical conventions

Clauses 10, 12, 20, 21, 22 and 30 are for description of cultural aspects in excess of what can be described in the corresponding POSIX clauses 1, 2, 3, 4, 5 and 6. In clause 9 it is possible to give further information on characters classified in clause 2.

NOTE: Further information about the categories, along with specific examples illustrating their use may be found in annex F and in the Nordic Cultural Requirements on Information Technology (Summary Report) (Annex G, item 1).

Each clause shall begin on a new line after at least one blank line, and each clause shall be introduced by the string "Clause ", followed by a decimal clause number, then a colon and a space, and then the title of the clause, using the titles above where applicable (examples are given in annex D).

The body of the clause shall follow on the succeeding lines. A reference to a clause within the specification shall consist of the string "=> Clause " followed by the clause number. A reference to another specification shall consist of the string "=> Spec. " followed by the registration number of the specification and, optionally, the string " Clause " and a clause number.

The format of the POSIX Locale and Charmap sources shall be conformant to ISO/IEC 9945-2, or for POSIX Locales the technique specified in Annex E.

The format of the POSIX Repertoiremap shall be conformant to 6.9.

6.3 The POSIX Locale shall define all standard categories (for example by copying categories of a standard POSIX Locale; examples are given in annex E).

6.4 The coded character set of ISO/IEC 646 International Reference Version (ISO 2375 registration number 6) shall be used to represent text for the submitted files. For enhanced network portability it is recommended that only the invariant part of ISO/IEC 646 (ISO 2375 registration number 170), which contains 83 graphical characters (including space), is used. Comments shall be given in the English language, and equivalent comments may also be given in other languages. If characters outside ISO/IEC 646 International Reference Version are needed, character names defined in a Repertoiremap shall be used.

6.5 The sources shall be delivered electronically, either via electronic mail or on a diskette to the Registration Authority. Narrative Cultural Specifications may alternately be delivered on paper.

6.6 A written application shall accompany the Cultural Specification and be signed by authorized personnel on behalf of the contributing organization. It shall release copyrights of the contributed sources.

6.7 The written Cultural Specification application shall contain information on the following items:

1. Cultural Specification type number (as in 6.2 above)
2. Organization name of Sponsoring Authority
3. Organization postal address
4. Name of contact person
5. Electronic mail address of the organization, or contact person
6. Telephone number for the organization, in international format.
7. Fax number for the organization, in international format.

For Types 1 and 2, Narrative Cultural Specifications and POSIX Locales:

8. Natural language, as specified in ISO 639
9. Territory, as two-letter form of ISO 3166

For Types 3 and 4, POSIX Charmaps and POSIX Repertoiremaps:

10. Suggested POSIX Charmap or POSIX Repertoiremap name

All applications shall contain information on these items:

11. If not for general use, an indication of the intended user audience. The Registration Authority decides on a corresponding identifier element, to be used in the token identifier for the specification.
12. If for use of a special application, a description of the application. The Registration Authority decides on a corresponding identifier element, to be used in the token identifier for the specification.
13. Short name for Sponsoring Authority, for possible use in the token identifier.
14. Revision number consisting of digits and zero or more full stops (".").
15. Revision date in the format according to this example: "1995-02-05" meaning the 5th of February, 1995.

If any of the above information is non-existent, it must be stated in each case; the corresponding string is then the empty string. The default case in 11 and 12 is also represented by an empty string. If required information is not present in ISO 639 or ISO 3166, the relevant Maintenance Authority shall be approached by the Sponsoring Authority to get the needed item registered.

The information in items 8 to 14 is used in the token identifier for the Cultural Specifications. Items 8 to 13 may contain digits 0123456789 and the characters uppercase and lowercase forms of

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ
```

Item 10 may also contain the special characters:

```
/()*-.:_
```

Case of letters is not significant in token identifiers.

Annex A specifies a form to be filled out for each Cultural Specification; Annex B gives an example of a filled-out form.

6.8 The information in item 8 to 14 is used by the Registration Authority to construct a token identifier for the Cultural Specification according to the following rules.

For Narrative Cultural Specifications and POSIX Locales the token identifier will be:

```
8_9+11+12,13_14
```

And for POSIX Charmaps and POSIX Repertoiremaps the token identifier will be:

```
10+11+12,13_14
```

where 11 and 12 and preceding pluses shall be omitted when not needed to specify position, and 13 may be omitted after request from the Sponsoring Authority.

NOTE: A combination of a POSIX Locale and a POSIX Charmap may be designated by the Locale identifier and the Charmap identifier separated by a solidus (/).

When referencing a Cultural Specification, the version number or parts thereof taken from the right may be omitted, to refer to the Cultural Specification with the highest digital version number available with the given version number prefix. If the item 13 is an empty string, referencing the token identifier without the preceding comma and items 13 and 14 shall give the Cultural Specification with the highest digital version number.

NOTE: The version number may be used by the Sponsoring Authority to mark major releases, minor revisions and error corrections. It is recommended that major releases be reflected as the first number, minor revisions in the second number, and error corrections in the third number.

EXAMPLE 1: EU,CEN_3.5 for the CEN European POSIX Locale
EXAMPLE 2: da_DK,_2.4 for the Danish Standards Danish POSIX Locale
EXAMPLE 3: ISO_8859-1:1987,DS_1.0 for the DS Charmap for ISO 8859-1

6.9 POSIX Locale and Charmap sources shall be specified in a way that is independent of coded character sets, using character names. Relation between the character names and characters shall be specified via a Repertoiremap table, giving the character name and the ISO/IEC 10646 canonical encoding in hexadecimal form (with a preceding 'U') and optionally the long ISO/IEC 10646 character name. It is recommended to use, whenever possible, character names specified in ISO/IEC 9945-2 Annex G. The character name and the ISO/IEC 10646 canonical encoding are each surrounded by angle brackets <>, and the fields shall be separated by one or more spaces or tabs on a line. If a right angle bracket or an escape character is used within a name, it shall be preceded by the escape character. The escape character can be redefined from the default reverse solidus (\) with the first line of the Repertoiremap containing the string "escape_char" followed by one or more spaces or tabs and then the escape character.

7 Appeal procedures

Appeal against the decision of the Registration Authority can be made as follows:

7.1 Appeal by a Sponsoring Authority can be made if it disagrees with the Registration Authority on whether the application meets the requirements of clause 6.

7.2 A Member Body of the JTC 1 subcommittee responsible for the maintaining of this International Standard may object to a forthcoming publication of a registration by the Registration Authority, but solely on the ground that the requirements in Clause 6 are not met.

7.3 Appeals shall be filed with the Registration Authority by registered mail

- either within 90 days of reception of the refusal of the Registration Authority

- or before the end of the circulation period.

7.4 Appeals shall be submitted by the Registration Authority within 90 days after reception in the case of 7.1 or at the end of the circulation period in the case of 7.2, to the members of the JTC 1 subcommittee responsible for the maintaining of this International Standard, to decide according to its voting procedures.

Annex A

(normative)

Application form for a Cultural Specification

Please specify all data relevant for the Cultural Specification type, indicating non-available data by "not available". Please fill out one form for each Cultural Specification submitted. When completed, please send it to the Registration Authority as listed in clause 4.

1. Cultural Specification type number: _____

2. Organization name of Sponsoring Authority: _____

3. Organization postal address: _____

_____

4. Name of contact person: _____

5. Electronic mail address of contact person: _____

6. Telephone number for contact person: + ___ _____

7. Fax number for contact person: + ___ _____

For Narrative Cultural Specifications and POSIX Locales (type 1 and 2):

8. Natural language, as specified in ISO 639: _____

9. Territory, as two-letter form of ISO 3166: _____

For POSIX Charmaps and POSIX Repertoiremaps (type 3 and 4):

10. The proposed POSIX Charmap or POSIX Repertoiremap name: _____

For all 4 types:

11. If not for general use, an intended user audience, e.g. librarians: _____

12. If for use of a special application, the short application name: _____

13. Short name for Sponsoring Authority, used in token identifier: _____

14. Version number with zero or more dots: _____

15. Revision date in ISO 8601 format: _____

The Cultural Specification identified above, and of which we hold copyright, is allowed for free distribution.

Date: _____ Authorized signature: _____

Annex B

(informative)

Sample Application for a Cultural Specification

Please specify all data relevant for the Cultural Specification type, indicating non-available data by "not available". Please fill out one form for each Cultural Specification submitted. When completed, please send it to the Registration Authority as listed in clause 4.

1. Cultural Specification type number: 1, Narrative Cultural Specification

2. Organization name of Sponsoring Authority: National Standards Authority of Ireland

3. Organization postal address: Glasnevin, Dublin 9, Ireland

4. Name of contact person: Seán Citizen

5. Electronic mail address of contact person: nsai@nsai.ie

6. Telephone number for contact person: +353 1 807-3800

7. Fax number for contact person: +353 1 807-3838

For Narrative Cultural Specifications and POSIX Locales (type 1 and 2):

8. Natural language, as specified in ISO 639: ga (Irish Gaelic)

9. Territory, as two-letter form of ISO 3166: IE (Ireland)

For POSIX Charmaps and POSIX Repertoiremaps (type 3 and 4):

10. The POSIX Charmap or POSIX Repertoiremap name: _____

For all four types:

11. If not for general use, an intended user audience, e.g. librarians: _____

12. If for use of a special application, short name of application: _____

13. Short name for Sponsoring Authority, used in token identifier: _____

14. Version number with zero or more dots: 0.5

15. Revision date in ISO 8601 format: 1996-01-28

The Cultural Specification identified above, and of which we hold copyright, is allowed for free distribution.

Date: 1996-03-16 Authorized signature: _____

Annex C

(normative)

External References to Cultural Specifications

C.1 Identification of Cultural Specifications

The Cultural Specifications registered according to this International Standard shall be referenced by object identifiers according to Abstract Syntax Notation 1 (ASN.1, ISO/IEC 8824 and ISO/IEC 8825).

C.2 Identification of Abstract Syntaxes

The definitions (abstract syntaxes) of the Cultural Specifications registered according to this International Standard shall be in the form of ASN.1 defined arcs which follow the arc which defines this International Standard.

The ASN.1 arc defining this International Standard is:

iso(1) identified-organization(3) ewos(16) eg(2) tlg(0) cultural-register(2)

Separate arcs shall be defined for type 1, 2, 3 and 4 specifications:

abstract-syntaxes(1) <cultural-type> <registration-number>

where <cultural-type> is the number of the type defined in clause 6, and <registration-number> is the numeric identifier assigned as per clause 4 responsibility g).

In the case of a type 2 Cultural Specification the standard categories as defined by the POSIX standard may be specified in additional arcs. This arc shall be category(1) and shall follow the arc specifying the registration
number:

abstract-syntaxes(1) <cultural-type> <registration-number> category(1) <locale-category>

The <locale-category> is a number as defined in 6.2 for the standard POSIX locale categories.

C.3 Object Descriptors

The object descriptors for the abstract syntax object identifiers defined in C.2 above shall be the name of this International Standard followed by a corresponding <identifier>, either numeric or token identifier, as assigned per clause 4 responsibility g):

ISO/IEC 15897 Cultural Specification Type <cultural-type> <identifier>

C.4 Transfer Syntax

The transfer syntax as specified in ISO 8824 defines the encoding in which the contents of a registry entry might be transferred over a network. For this purpose the transfer syntaxes as defined in ISO/IEC 2022 shall be used.

Annex D

(informative)

Sample Narrative Cultural Specifications for Danish and Irish

A description of the clauses can be found in annex F.

D.1. Danish language locale for Denmark, Narrative Cultural Specification

Users: general, applications: general
Source: Dansk Standard, date: 1994-07-28, version: 2.4
Token identifier: da_DK,_2.4

Clause 1: Alphanumeric deterministic ordering

Ordering in Danish is defined in Danish Standard DS 377, 3rd edition (1980) and the Danish Orthography Dictionary ("Retskrivningsordbogen", Gyldendal, København 1986. ISBN 87-00-91372-3).

Normal <a> to <z> ordering is used on the Latin script, except for the following letters: The letters <æ> <ø> <å> are ordered as 3 separate letters after <z>. <ü> is ordered as <y>, <ä> as <æ>, <ö> as <ø>, <ð> as <d>, <þ> as <t><h>, French <œ> as <o><e>. Two <a>s are ordered as <å>, except when denoting two sounds (which is normally the case only in combined words). Nonaccented letters come before accented letters, and capital letters come before small letters, when words otherwise compare equally. There is no explicit ordering of accents specified in "Retskrivningsordbogen", and whether case or accents are the most important is not specified.

Clause 2: Classification of characters

Danish uses normal classification of letters in uppercase and lowercase, this classification is also applicable to scripts like Greek and Cyrillic.

Clause 3: Numeric formatting

The decimal separator is COMMA <,>
The thousands separator is FULL STOP <.>
The grouping of large numbers is in groups of three digits.

Clause 4: Monetary formatting

International currency symbol:	DKK 543,21
Domestic currency symbol:	kr 543,21
Use of negative sign:	kr -543,21
Thousands and decimal separators:	kr 9.876.543,21

Clause 5: Date and time conventions

Both weekday and month names are written with an initial lowercase letter in Danish (Normal capitalizing rules apply in the beginning of a sentence, etc.).

English name	Weekday names	Short weekday names
Sunday	søndag	søn
Monday	mandag	man
Tuesday	tirsdag	tir
Wednesday	onsdag	ons
Thursday	torsdag	tor
Friday	fredag	fre
Saturday	lørdag	lør

Short weekday names consisting of the two first letters are also commonly used.

English name	Month name	Short month name
January	januar	jan
February	februar	feb
March	marts	mar
April	april	apr
June	juni	jun
July	juli	jul
August	august	aug
September	september	sep
October	oktober	okt
November	november	nov
December	december	dec

Long date:	07 juni 1994
Abbreviated day and time:	tir 07 jun 1994 23:22:33 CET DST
long date with weekday:	onsdag den 21. december 1994
Abbreviated long date:	07 jun 1994
Numeric date:	1994-06-07
Time:	18:06:20

The 24 hour system is used in Denmark. There are no abbreviations commonly in use for before or after noon.

Clause 6: Affirmative and negative answers

Yes expressions	1JjYy	(= 1, Ja, Yes)
No expressions	0Nn	(= 0, Nej, No)

Clause 7: National or cultural Information Technology terminology

The official Information Technology terminology is "Edb-ordbog", DS 2049-1970, Gjellerup, København. A newer description can be found in Lars Frank: "edb-ordbogen", Kommunetryk, København 1984.

Clause 8: National or cultural profiles of standards

POSIX ISO/IEC 9945-1:1990 annex F and ISO/IEC 9945-2:1993 annex G contains example Danish POSIX profiles.

Clause 9: Character set considerations

The following is the Danish alphabet:

Aa Bb Cc Dd Ee Ff Gg Hh Ii Jj Kk Ll Mm Nn Oo Pp Qq Rr Ss Tt Uu Vv Ww Xx Yy Zz Ææ Øø Åå

The combination of two <a>s are regarded as one <å>, originating from older orthography but still used in many person and place names.

For indicating stress, different pronunciation and long vowels, an accent can be used on all vowels:

Áá Éé ÍÍ Óó Úú Ýý Ǽǽ Ǿǿ Åå

The following letters of foreign origin is commonly used in Danish newspapers and books, according to examples in "Retskrivningsordbogen":

Ää Àà Çç Ðð Êê Ëë Èè Ôô Őő Œœ Þþ Üü Ää Öö

The recommended character set is DS/ISO 8859-1; for a bigger repertoire DS/ISO/IEC 10646-1 is recommended.

Other character standards in use include ISO/IEC 6937 and ISO/IEC 646 (a Danish version, DS 2089, of this has been withdrawn, but is still in use).

Vendor character sets in use include HP Roman 8, IBM CP 277, 278, 437, 850, 865, Macintosh, and MS CP 1252.

The character sets have been described in the Internet RFC 1345, made in a Danish Standards Association and INSTA project, and they are also available in POSIX Charmap format.

Danish Internet Email exchange recommends MIME format and ISO 8859-1 encoding, if necessary in RFC1345 mnemonic format.

The Danish EDI council recommends ISO 8859-1 as the exchange coded character set, with possible RFC1345 mnemonic extensions.

Clause 10: Sorting and searching rules

The character oriented ordering is described in => Clause 1. More sophisticated ordering as described in "Retskrivningsordbogen" requires that numbers are expanded to their spelling, and also special characters be expanded to their spelling before sorting. Also a number of common words are to be discarded before sorting, such as "den", "det", "en", "et".

Clause 11: Transformation of characters

Transliteration of Cyrillic and Arabic is very different from English conventions.

For a fallback notation of some letters, refer to the following table:

original letter	2-char	1-char
Æ	AE	E
Ø	OE	Y
Å	AA	O
Ü	Y	Y
Ä	Æ	Æ
Ö	Ø	Ø
Þ	TH	T

Clause 12: Character properties

For ordinary classification of characters, please refer to => Clause 2.

The Greenlandic letter KRA <ĸ> has no uppercase equivalent, and is converted to a "Q" as also prescribed by modern Greenlandic orthography.

Clause 13: Use of special characters

For quoting, the characters <"><">, <»><«> and <“><”> are used, with the shown order.

Various punctuation signs:

NUMBER SIGN <#> is seldomly used, and should be avoided

AT SIGN <@> is not used for commercial purposes. It is used in Internet mail.

Double space after a FULL STOP <.> is not used.

DIVISION SIGN <÷> should not be used for division, as it is also used for subtraction, the sign is known as "minus" in Denmark. Use SOLIDUS </> instead.

SECTION SIGN <§> is often used in legal documents to refer to paragraphs.

In a sentence the FULL STOP <.> is placed as the last character, as in: Skibet hed "Titanic".

Clause 14: Character rendition

The Danish letters <Ø> and <ø> are often misprinted. The stroke in the letters is the problem. If you consider a rectangle box surrounding the letter, then the stroke should cross from the upper right corner to the opposite corner.

Clause 15: Character inputting

A proposed general input method is included in DS/ISO/IEC 9945-1 annex F.

Clause 16: Personal names rules

Children can get their father's or mother's last name, or any combination of these with or without a hyphen. Also in marriage the bride and the groom may take the other partner's name in any combination.

Personal names are commonly spelt with the full first name, while use of initials only is seen also. People are mostly addressed by voice by their first name. The common address form is the informal "du", and the more formal "De" is becoming more common. The family name is never spelt in capital letters only, contrary to continental European habits. Titles are used in some circumstances.

Clause 17: Inflection

The Danish grammar is defined in "Retskrivningsordbogen". Danish has more inflections than English, for example nouns will have 8 forms based on indefinite/definite, singularis/pluralis and nominative+others/genitive.

Danish tends to have longer words then English, as you can make combined words.

Clause 18: Hyphenation

Hyphenation rules are described in "Retskrivningsordbogen".

Clause 19: Spelling

Spelling of the Danish language is specified in "Retskrivningsordbogen". This spelling is approved by Danish Government, and used as authoritative in schools etc.

Clause 20: Numbering, ordinals and measuring systems

See => Clause 3 and => Clause 4 for a description of numeric and monetary formatting.

The measurement system is the SI system, DS/ISO 1000.

Temperatures are normally measured in degrees Celsius, the Kelvin scale is sometimes used in science.

Clause 21: Monetary amounts

See => Clause 4 for the POSIX specifications.

Clause 22: Date and time

The timezone is UTC+0100 in the winter, UTC+0200 in the summer. The daylight savings period currently (1995) changes by one hour the last Sunday in March at 02:00, and back again by one hour the last Sunday in September at 03:00. This may change in the future. There is no official names for the timezones.

Use of week numbers are very common, and the week numbering is according to DS/ISO 8601.

The first day of the week is Monday, in accordance with DS/ISO 8601.

Date formatting according to DS/ISO 8601, for example 1995-04-13 for 13th of April 1995, is very common in technical business and in legal business, and other areas.

For POSIX date and time formatting, please see => Clause 5.

Clause 23: Coding of national entities

Denmark is a part of the Kingdom of Denmark, which also consists of Greenland and the Faroe Islands.

Denmark is situated about 54 - 58 degrees North, and 8 - 15 degrees East.
Denmark has an area of about 43.069 km2 and 5,2 mill inhabitants (1995).
The main language is Danish.

There are a number of standards giving a country code to Denmark:

ISO 3166 alpha-2	DK
ISO 3166 alpha-3	DNK
ISO 3166 numeric	208
CEPT-MAILCODE	DK
UN Genève 1949:68 Vehicle code	DK
CCITT E.163 international telephone prefix	45
CCITT X.121 X.25 numbering country code	238
ISO 2108 ISBN book numbering	87

The Alpha-2 code "DK" of ISO 3166 is for general use, and is use generally by the public as the abbreviation for Denmark.

The name of the country in Danish is "Danmark".

The language code according to ISO 639 for the Danish language is "da".

The name of the Danish language in Danish is "dansk".

The currency is Danish Kroner, in Danish, "danske kroner". The ISO 4217 code is DKK. The native abbreviation is "kr". 1 "krone" is equal to 100 "øre". See => Clause 4 for a POSIX description.

Postal codes ("postnumre") are 4 digits. See => Clause 25 for their use.

For public adminstration Denmark has 14 counties ("amter") and 275 communes ("kommuner"). The counties and communes have numbers, which can be found in Statistic Yearbook from Denmark’s Statistics.

Clause 24: Telephone numbers

The international telephone prefix for Denmark is +45. There are no area codes; all numbers have 8 digits.
The recommended format for telephone numbers is in groups of 2, for example 39 17 99 44.

Clause 25: Mail addresses

See => Clause 16 for how to write personal names.

The street number is placed after the street name.

The postal code is placed before the city name. The CEPT country prefix should be placed in front of the postal code for international mail, this is even commonly done for mail within Denmark.
Postal codes are defined in "Post- og Telegrafhåndbogen - Postnummerdelen", obtainable at all postal offices, and may be found also in telephone directory books.

An example of a mail address is:

Danish Standards Association
Att: S142 u22 A8
Baunegaardsvej 73
DK-2900 Hellerup
Danmark

According to CEPT recommendations, one should either use the French name of the country ("Danemark"), or the name in the local language "Danmark".

Storey specification is placed after the street number. The following conventions apply:

English	Danish	Danish abbreviation
Ground floor	stuen	st
1st floor	1. etage	1
basement	kælderen	kld
right	til højre	th
left	til venstre	tv
middle	midt for	mf

An example of its use:

Holger Danske
Fremtidsvej 26, 2 tv
DK-2000 Frederiksberg

Clause 26: Identification of persons and organizations

In Denmark, persons are identified by a unique personal identity number ("personnummer" or "CPR-nummer"). This number incorporates the date of birth and the sex. The structure of the Danish personal identity number is:

DDMMYY-XNNB

where DD=day, MM=month, YY=year, X=running number, including century indication, NN=running number, and B=running number with sex indication: odd=male; even=female.

Danish organisations are identified via the SE-number, which is also used as an identification for Value Added Tax (VAT) purposes. This is an 8-digit number, the VAT number may have a 2-digit area code attached in the end, after a hyphen.

There are a number of official registries for organizations, dependent of the organisation form: "Aktieselskabsregistret", "Anpartselskabsregistret", "Fondsregistret".

Clause 27: Electronic mail addresses

The Danish X.400 email country code is DK, that is the ISO 3166 alpha-2 code.

The Danish Internet top domain is .DK (ISO 3166 alpha-2 code). Internet domain addresses have an organization name as the second level name. There are no economic sector (government, commercial, academic) indication.

The Danish X.500 service uses the character set T.61 with RFC 1345 mnemonic extensions for representing names and addresses.

Clause 28: Payment account numbers

The format of Danish bank account numbers have a 4-digit branch identification code, and then the numeric bank account number.

The format of the Danish Postal Giro accounts is 7 digits, an example is 123-4567.

Clause 29: Keyboard layout

A Danish keyboard has the layout of the alphabetic keys (first is lowercase, second is uppercase, third is alternate graphic):

```
½§  1!  2"@ 3#£ 4¤$ 5%  6&  7/{ 8([ 9)] 0=} +?  ´‘|
     Q   W   E   R   T   Y   U   I   O   P   Å   ¨^~
      A   S   D   F   G   H   J   K   L   Æ   Ø   ’*
   <>\ Z   X   C   V   B   N   M  ,;  .:  -_
```

´‘¨^~ are normally dead keys.

Clause 30: Man-machine dialogue

Naturally, most Danish users require programs where all menus, names of icons, commands, information messages, help texts, manuals etc. are translated and adjusted to their language and culture.

Programmers and screen layout designers must bear in mind that when English text is translated into Danish - and most other languages - it will normally be longer, i.e. require more space on the screen and occupy more computer memory.

Denmark has its own cultural symbols in some cases and use of non-Danish symbols as icons can create irritation and - if they are not easily recognized - confusion. Example: The typical suburban American mailbox with the raised flag is unusual in Denmark and hence not immediately associated with mail for most users.

Clause 31: Paper formats

ISO 216 paper sizes are used in Denmark. Two holes or 4 holes according to ISO 838 in A4 paper etc. is very common.

Clause 32: Typographical conventions

In Denmark the Didot point measure is used in typography, which is 7% larger than the point used in English and American typography.

When justifying text at both margins, extra space should be inserted between words, not between letters within a word.

Use of special characters are described in => Clause 13.
End of Narrative Cultural Specification.

D.2 Irish language locale for Ireland, Narrative Cultural Specification

Irish language locale for Ireland
Users: general
Applications: general
Narrative Cultural Specification
Source: National Standards Authority of Ireland, version: 0.5 (Unofficial draft)
Token identifier: ga_IE,_0.5

Timpeallacht teanga Ghaeilge na hÉireann.
Úsáideoirí: úsáideoirí i gcoitinne
Feidhmeanna: feidhmeanna i gcoitinne
Prósaithris ar Mhionsonraí Cultúrtha
Foinse: Údarás Náisiúnta Caighdeán na hÉireann, leagan: 0.5 (Dréacht neamhoifigiúil)
Comhartha aitheantais: ga_IE,_0.5

Clause 1: Alphanumeric deterministic ordering

Modern Irish Gaelic is written with the Latin alphabet, ordered from A to Z. Accented characters are ordered as their unaccented counterparts. In the case of equality after this comparison has been made, unaccented characters are ordered before accented characters, and capital letters are ordered before small letters.

Cuid 1: Rangú docht alfa-uimhriúil

Is iondúil go n-úsáidtear aibítir na Laidine chun Ghaeilge an lae inniu a scríobh. Déantar na litreacha sin a rangú in ord ó A go Z. Déantar litreacha ar a bhfuil comharthaí idirdhealaitheacha a rangú díreach ar nós na coda eile. Tugtar tús áite don litir lom laistigh den rangú sin, agus tugtar tús áite laistigh den rangú sin do litreacha móra thar litreacha beaga.

Clause 2: Classification of characters

Irish Gaelic classifies Latin letters as capital and small letters, and employs two different kinds of Latin script. In the Roman variant of the Latin script, each small letter corresponds to one capital letter. In the Gaelic variant of the Latin script, two letters have two forms each (ſr, ʀꝛ) corresponding to one capital letter (S, ʀ).

Cuid 2: Aicmí carachtar

Déantar idirdhealú sa Ghaeilge idir litreacha móra agus litreacha beaga na Laidine, agus úsáidtear dá chineál cló. Sa chló Rómhánach, tá aon litir bheag amháin ag freagairt don aon litir mhór amháin. Sa chló Ghaelach, tá dhá litir a bhfuil dhá fhoirm éagsúla díobh (ſr, ʀꝛ) ag freagairt don aon litir mhór amháin (S, ʀ).

Clause 3: Number formatting

The decimal separator is FULL STOP "."
The thousands separator is COMMA ","
The grouping of large numbers is in groups of three digits.

Cuid 3: Formáidí uimhriúla

LÁNSTAD atá mar dheighilteog dheachúil "."
CAMÓG atá mar dheighilteog idir mílte ","
Roinntear uimhreacha móra i ngrúpaí trí charachtar.

Clause 4: Monetary formatting

International currency symbol	IEP 543.21
Domestic currency symbol	£543.21
Use of negative sign	-£543.21
Thousands and decimal separators:	£9,876,543.21

Sometimes IR£543.21 is used for the domestic currency symbol.

Cuid 4: Formáidí airgid

Comhartha idirnáisiúnta airgeadais	IEP 543.21
Comhartha inmheánach airgeadais	£553.21
Suíomh an comhartha diúltach	-£543.21
Deighilteog dheachúil agus deighilteog idir mílte:	£9,876,543.21

Uaireanta úsáidtear an fhoirmle IR£543.21 mar chomhartha inmheánach airgeadais.

Clause 5: Date and time conventions

Names of days and names of months are written with an initial capital letter in Irish.

English name	*Weekday name*	*Short name*
Sunday	Dé Domhnaigh	Domh
Monday	Dé Luain	Luan
Tuesday	Dé Máirt	Máir
Wednesday	Dé Céadaoin	Céad
Thursday	Déardaoin	Déar

Cuid 5: Formáidí ama agus dáta

Tosaíonn ainmneacha laethe agus ainmneacha na míonna le litreacha móra sa Ghaeilge.

Béarla	*Lá*	*Ainm giorraithe*
Sunday	Dé Domhnaigh	Domh
Monday	Dé Luain	Luan
Tuesday	Dé Máirt	Máir
Wednesday	Dé Céadaoin	Céad
Thursday	Déardaoin	Déar

Friday	Dé hAoine	Aoin
Saturday	Dé Sathairn	Sath

English name	*Month name*	*Short name*
January	Eanáir	Ean
February	Feabhra	Feabh
March	Márta	Már
April	Aibreán	Aib
May	Mí na Bealtaine	Beal
June	Meitheamh	Meith
July	Iúil	Iúil
August	Lúnasa	Lún
September	Meán Fómhair	MFómh
October	Deireadh Fómhair	DFómh
November	Mí na Samhna	Samh
December	Mí na Nollag	Noll

Long date with day and night time:
Céadaoin, 21 Mí na Nollag 1994, 23:22:33 GMT
Long date with day and day time:
Dé Máirt, 7 Meitheamh 1994, 11:22:33 GMT
Abbreviated long date with night time:
Céad 21 Noll 1994 23:22:33
Abbreviated long date with day:
Máir 7 Meith 1994 11:22:33
Long date: 21 Mí na Nollag 1994
Long date: 7 Meitheamh 1994
Abbreviated long date: 21 Noll 1994
Abbreviated long date: 7 Meith 1994
Short date: 1994-12-21
Short date: 1994-06-07

For night time, 11:22:33 i.n. (*iarnóin*) is sometimes used. For day times, 11:22:33 r.n. (*roimh noin*) is sometimes used. The forms 21/12/94, 7/6/94 are also used. The month names *Bealtaine*, *Samhain*, and *Nollaig* are often used, but as these refer to individual days of the month (1 May, 1 November, 25 December), this use is not recommended.

Clause 6: Affirmative and negative answers

Yes expressions	TtSsYy1+	(= Tá, Sea, Yes)
No expressions	NnCc0-	(= Níl, Chan)

Clause 9: Character set considerations

The traditional Irish Latin alphabet consists of eighteen letters (*a b c d e f g h i l m n o p r s t u*). At the present time, the following letters of the Latin alphabet are used:

Aa Áá Bb Cc Dd Ee Éé Ff
Gg Hh Ii Íí Jj Kk Ll Mm Nn Oo Óó
Pp Qq Rr Ss Tt Uu Úú Vv
Ww Xx Yy Zz

Friday	Dé hAoine	Aoin
Saturday	Dé Sathairn	Sath

Béarla	*Mí*	*Ainm giorraithe*
January	Eanáir	Ean
February	Feabhra	Feabh
March	Márta	Már
April	Aibreán	Aib
May	Mí na Bealtaine	Beal
June	Meitheamh	Meith
July	Iúil	Iúil
August	Lúnasa	Lún
September	Meán Fómhair	MFómh
October	Deireadh Fómhair	DFómh
November	Mí na Samhna	Samh
December	Mí na Nollag	Noll

Dáta fada agus tráth den oíche:
Céadaoin, 21 Mí na Nollag 1994, 23:22:33 GMT
Dáta fada agus tráth den lá:
Dé Máirt, 7 Meitheamh 1994, 11:22:33 GMT
Dáta fada giorraithe agus tráth den oíche:
Céad 21 Noll 1994 23:22:33
Dáta fada giorraithe agus tráth den lá:
Máir 7 Meith 1994 11:22:33
Dáta fada: 21 Mí na Nollag 1994
Dáta fada: 7 Meitheamh 1994
Dáta fada giorraithe: 21 Noll 1994
Dáta fada giorraithe: 7 Meith 1994
Dáta gearr: 1994-12-21
Dáta gearr: 1994-06-07

Úsáidtear an fhoirmle 11.22.23 i.n. ar uairibh, ag tagairt don tráth den oíche, agus 11:22:23 r.n. ag tagairt don tráth den lá. Úsáidtear 21/12/94, 7/6/94 chomh maith. Úsáidtear *Bealtaine*, *Samhain*, agus *Nollaig* ag tagairt do na míonna go minic, ach ón uair gur laethe ar leith atá i gceist sna focail sin ó cheart, is í an fhoirmle *Mí na Bealtaine*, *Mí na Samhna*, agus *Mí na Nollag* a mholtar.

Cuid 6: Freagraí dearfacha agus freagraí diúltacha

Freagra dearfach	TtSsYy1+	(= Tá, Sea, Yes)
Freagra diúltach	NnCc0-	(= Níl, Chan)

Cuid 9: Tacair carachtar

Ocht litir dhéag (*a b c d e f g h i l m n o p r s t u*) atá in aibítir thraidisiúnta Laidine na Gaeilge. Is iad seo a leanas na litreacha den aibítir Laidine atá in úsáid inniu:

Aa Áá Bb Cc Dd Ee Éé Ff
Gg Hh Ii Íí Jj Kk Ll Mm Nn Oo Óó
Pp Qq Rr Ss Tt Uu Úú Vv
Ww Xx Yy Zz

The Latin letters JKQWXYZ are commonly found in loan words and personal names. Prior to the middle of this century, the Gaelic variant of the Latin script was more often used in print than the Roman variant; it is still in use in handwriting.

Aa Áá Bb Bḃ Cc Cċ Dd Dḋ Ee Éé Ff Fḟ
Gg Gġ Hh Ii Íí Jj Kk Ll Mm Mṁ Nn Oo Óó
Pp Pṗ Qq Rr Ss Sṡ Tt Tṫ Uu Úú Vv
Ww Xx Yy Zz

Two diacritical marks are in use: an acute accent may occur on any vowel, and a dot may be written as a mark of lenition on nine of the consonants (*b c d f g m p s t*). The consonants traditionally written with dots as a mark of lenition are usually written today with a following *h* instead (ḃ = *bh*).

The Scottish Gaelic letters *Àà Èè Ìì Òò Ùù* are sometimes found in Irish texts.

Clause 16: Personal names rules

Irish surnames commonly begin with *Mac, Ó, de, Ní, Nic* or *Uí*. For sorting purposes these should be ignored, and surnames sorted by the next element in them, so that people who have the same surname (Pádraig Ó Néill, Máire Uí Néill, for example) are grouped together.

Clause 17: Inflection

Irish Gaelic is a highly inflected language, in which the beginning, middle, and end of words may change, according to rules of grammar. A full account of the rules under which lenition, eclipsis and other changes occur is found in the Christian Brothers' *New Irish grammar* (Dublin: CJ Fallon 1990), or in Irish in *Graiméar Gaeilge na mBráithre Críostaí*, (Baile Átha Cliath: MH Mac an Ghoill agus a Mhac, 1960).

Is in iasachtaí déanacha agus in ainmeacha dilse amháin a úsáidtear na litreacha Laidine JKQWXYZ. Go dtí lár na haoise seo, ba é an cló Gaelach ba mhó a bhí in úsáid i saothar foilsithe, agus tá sé in úsáid i gconaí i saothar lámhscríofa.

Aa Áá Bb Bḃ Cc Cċ Dd Dḋ Ee Éé Ff Fḟ
Gg Gġ Hh Ii Íí Jj Kk Ll Mm Mṁ Nn Oo Óó
Pp Pṗ Qq Rr Ss Sṡ Tt Tṫ Uu Úú Vv
Ww Xx Yy Zz

Dhá chomhartha idirdhealaitheacha atá in úsáid: is féidir agúid a chur ar ghuta agus ponc séimhithe ar na naoi gconsan seo a leanas: *b, c, d, f, g, m, p, s, t*. Is é nós an lae inniu é an litir *h* a chur i ndiaidh consain, mar chomhartha séimhithe ar an gconsan sin, in áit ponc a chur os a chionn (ḃ = *bh*).

Is minic a gheibhtear na litreacha Albanacha *Àà Èè Ìì Òò Ùù* i dteacsanna Gaeilge.

Cuid 16: Ordú sloinnte

Is iondúil go dtosaíonn sloinnte in Éirinn le ceann de na focail seo: *Ó, Mac, de, Ní, Nic* nó *Uí*. Is ceart sloinnte a chur in ord de réir na chéad mhíre eile díobh, chun daoine a bhfuil an sloinne céanna acu (Pádraig Ó Néill, Máire Uí Néill, mar shampla) a rangú le chéile.

Cuid 17: Claochlú focal

Is minic a athraíonn tús, lár, nó deireadh focail de réir gramadaí. Tá cur síos mion le fáil ar na rialacha a ghabhann le séimhiú, le hurú, agus le hathruithe eile le fáil in *Graiméar Gaeilge na mBráithre Críostaí*, (Baile Átha Cliath: MH Mac an Ghoill agus a Mhac, 1960), nó i mBéarla in *New Irish grammar* (Dublin: CJ Fallon 1990).

Annex E

(normative)

"reorder-after" construct in POSIX LC_COLLATE

For the LC_COLLATE definition in POSIX, additional keywords "reorder-after" and "reorder-end" are allowed when the "copy" keyword is used. This changes the specification for the "copy" keyword, as other keywords are then allowed. The "copy" shall refer a collating specification in source form not using the "copy" keyword.

The following keywords are recognized in a collating sequence definition, in addition to the keywords specified in ISO/IEC 9945-2:

reorder-after Redefine collating rules. Specify after which collating element the redefinition of collation order shall take order. This statement is followed by one or more collation order statements, reassigning character collation values and collation weights to collating elements.

reorder-end Specify the end of the "reorder-after" collating order statements.

E.1 "reorder-after" keyword

The "reorder-after" keyword shall be used to specify a modification to a copied collation specification of an existing locale. There can be more than one "reorder-after" statement in a collating specification. Using ISO/IEC 9945-2 specification methods, the syntax shall be:

"reorder-after %s\n",<collating-symbol>

The <collating-symbol> operand shall be a name, enclosed between angle brackets, and shall be present in the source locale copied via the "copy" keyword.

The "reorder-after" statement is followed by one or more collation identifier entries as described in the "Collating Order" clause (2.5.2.2.4 in ISO/IEC 9945-2), with the exception that the ellipsis symbol (...) shall not be used.

Each collation identifier entry reassigns character collation values and collation weights to collating elements existing in the copied collation specification, by removing the collating identifier entry from the copied specification, and inserting the collating element in the collating sequence with the new collation weights after the preceding collating element of the "reorder-after" specification, the first collating element in the collation sequence being the <collating-symbol> specified on the "reorder-after" statement.

A "reorder-after" specification is terminated by another "reorder-after" specification or the "reorder-end" statement.

E.2 "reorder-end" keyword

The "reorder-end" keyword shall specify the end of a list of collating identifier entries, initiated by the "reorder-after" keyword.

E.3 Example of "reorder-after": (this clause is informative)

```
reorder-after <y8>
<U:>        <Y>;<U:>;<CAPITAL>
<u:>        <Y>;<U:>;<SMALL>
reorder-after <z8>
<AE>        <AE>;<NONE>;<CAPITAL>
<ae>        <AE>;<NONE>;<SMALL>
<A:>        <AE>;<DIAERESIS>;<CAPITAL>
<a:>        <AE>;<DIAERESIS>;<SMALL>
<O/>        <O/>;<NONE>;<CAPITAL>
<o/>        <O/>;<NONE>;<SMALL>
<AA>        <AA>;<NONE>;<CAPITAL>
<aa>        <AA>;<NONE>;<SMALL>
reorder-end
```

The example is interpreted as follows:

1. The collating element <U:> is removed from the copied collating sequence and inserted after <y8> in the collating sequence with the new weights. The collating element <u:> is removed from the copied collating sequence and inserted in the resulting collation sequence after <U:> with the new weights.

2. The second "reorder-after" statement terminates the first list of reordering collation identifier entries, and initiates a second list, rearranging the order and weights for the <AE>, <ae>, <A:>, <a:>, <O/>, and <o/> collating elements after the <z8> collating symbol in the copied specification.

3. The "reorder-end" statement terminates the second list of reordering entries.

4. Thus for the original sequence

   ... (Uu Üü) Vv Ww Xx Yy Zz

   this example reordering gives

   ... Uu Vv Ww Xx (Yy Üü) Zz (Ææ Ää) Øø Åå

5. A complete example for Danish is included in E.6.1. For the sequence

   ... (Uu Ùù Úú Üü Űű) Vv Ww Xx (Yy Ýý) (Zz Žž)

   the example reordering in E.6.1 gives

   ... (Uu Ùù Úú) Vv Ww Xx (Yy Ýý Üü Űű) (Zz Žž)

   (Ææ Ǽǽ Ǣǣ Ää) (Øø Ǿǿ Öö Őő) (Åå (AA Aa aA aa) Åå)

E.4 "reorder-after" rationale (this clause is informative)

Much work has already been done on locales, and making them quite general. CEN/TC304 has on its programme of work the specification of a multilingual ordering for Europe, and also the collection of collating sequences of the different countries in Europe in a registry of cultural elements; one of the formal specification techniques for this is using the POSIX standard (ISO/IEC 9945-2:1993). The POSIX standard introduces a copy command for all sections of the POSIX locale. This is good for many purposes and it ensures that two locales are equivalent for this category. A further step in building on previous locale work is defined in this International Standard.

Collating sequences often vary a bit from country to country, and from language to language, but generally much of the collating sequence is the same. For instance the Danish sequence is for the most part the same as the German, English or French collation, but for about a dozen letters it differs. The same can be said for Swedish or Hungarian: generally the Latin collating sequence is the same, but a few characters are different.

With the advent of the quite general, coded character set independent locales like the example Danish locales in the POSIX Shell and Utilities standard (ISO/IEC 9945-2:1993) annex G, and the European multilingual POSIX collating specification, it will prove to be convenient if the few differences could be specified just as changes to an existing one. Using the "reorder-after" construct will also help improve the overview of what the changes really are for implementers and other users.

An example of the use of the "reorder-after" construct is the following. A default European ordering for the Latin alphabet may be adequate for Danish, with the exception of the collation rules for the letters Ü, ü, Æ, æ, Ä, ä, Ø, ø, Å and å. By applying the "reorder-after" construct, the Danish specification can be made more easily by copying and reordering the existing European specification, rather than specifying collation parameters for all Latin letters (with or without diacritics). There is no obligation for Denmark to take this approach, but the normative annex E provides the mechanism for doing so if it is deemed desirable.

E.5 awk script for "reorder-after" construct (this clause is informative)

A script has been written in the "awk" language defined in the POSIX standard ISO/IEC 9945-2 to implement the "reorder-after" construct:

```
BEGIN { comment = "%"; back[0]= follow[0] = 0; }
/LC_COLLATE/ { coll=1 }
/END LC_COLLATE/ { coll=0; for (lnr= 1; lnr; lnr= follow[lnr]) print c-
ont[lnr] }

{ if (coll == 0) print $0  ;
   else { if ($1 == "copy")   {
                file = $2
                while (getline < file )
                if ( $1 == "LC_COLLATE" ) copy_lc = 1
                else if ( $1 == "END" && $2 == "LC_COLLATE" ) copy_lc =0
                else if (copy_lc) {
                                lnr++
                                follow[lnr-1] = lnr; back [ lnr ] = lnr-1
                                cont[lnr] = $0; symb[ $1 ] = lnr
                }
                close (file )
     }
     else if ($1 == "reorder-after") { ra=1 ; after = symb [ $2 ] }
     else if ($1 == "reorder-end") ra = 0
     else {
                lnr++
                if (ra) follow [ lnr ] = follow [ after ]
                if (ra) back [ follow [ after ] ] = lnr
                follow[after] = lnr; back [ lnr ] = after
                cont[lnr] = $0
                if ( ra && $1 != comment && $1 != "" )  {
                                old = symb [ $1 ];
                                follow [ back [ old ] ] = follow [ old ];
                                back [ follow [ old ] ] = back [ old ];
                                symb[ $1 ] = lnr;
                }
                after = lnr
     }
   }
}
```

E.6 Sample POSIX Locale Specifications for Danish and Irish Gaelic (this clause is informative)

E.6.1 Danish example

```
escape_char /
comment_char %

% Danish language locale for Denmark
% Source: Danish Standards Association
% Address: Baunegaardsvej 73,
%   DK-2900 Hellerup, Danmark
% Contact: Keld Simonsen
% Email: Keld.Simonsen@dkuug.dk
% Tel: +45 - 39770101
% Fax: +45 - 39770202
% Language: da
% Territory: DK
% Revision: 4.0
% Date: 1995-04-12
% Application: general
% Users: general
% Repertoiremap: mnemonic,DS
% Charset: ISO_8859-1:1987
% Distribution and use is free, also
% for commercial purposes.

% The ordering algorithm is in accordance
% with Danish Standard DS 377 (1980)
% and the Danish Orthography Dictionary
% (Retskrivningsordbogen, 1986).
% It is also in accordance with
% Greenlandic orthography.

LC_COLLATE
collating-element <A-A> from "<A><A>"
collating-element <A-a> from "<A><a>"
collating-element <a-A> from "<a><A>"
collating-element <a-a> from "<a><a>"
copy en_DK
reorder-after <CAPITAL>
<CAPITAL>
<CAPITAL-SMALL>
<SMALL-CAPITAL>
<SMALL>
reorder-after <q8>
<kk>            <Q>;<SPECIAL>;<SMALL>;IGNORE
reorder-after <t8>
<TH>            "<T><H>";"<TH><TH>";"<CAPITAL><CAPITAL>";IGNORE
<th>            "<T><H>";"<TH><TH>";"<SMALL><SMALL>";IGNORE
reorder-after <y8>
% <U:> and <U"> are treated as <Y> in Danish
<U:>            <Y>;<U:>;<CAPITAL>;IGNORE
<u:>            <Y>;<U:>;<SMALL>;IGNORE
<U">            <Y>;<U">;<CAPITAL>;IGNORE
<u">            <Y>;<U">;<SMALL>;IGNORE
reorder-after <z8>
% <AE> is a separate letter in Danish
<AE>            <AE>;<NONE>;<CAPITAL>;IGNORE
<ae>            <AE>;<NONE>;<SMALL>;IGNORE
<AE'>           <AE>;<ACUTE>;<CAPITAL>;IGNORE
<ae'>           <AE>;<ACUTE>;<SMALL>;IGNORE
<A3>            <AE>;<MACRON>;<CAPITAL>;IGNORE
<a3>            <AE>;<MACRON>;<SMALL>;IGNORE
<A:>            <AE>;<SPECIAL>;<CAPITAL>;IGNORE
<a:>            <AE>;<SPECIAL>;<SMALL>;IGNORE
% <O//> is a separate letter in Danish
```

```
<O//>           <O//>;<NONE>;<CAPITAL>;IGNORE
<o//>           <O//>;<NONE>;<SMALL>;IGNORE
<O//'>          <O//>;<ACUTE>;<CAPITAL>;IGNORE
<o//'>          <O//>;<ACUTE>;<SMALL>;IGNORE
<O:>            <O//>;<DIAERESIS>;<CAPITAL>;IGNORE
<o:>            <O//>;<DIAERESIS>;<SMALL>;IGNORE
<O">            <O//>;<DOUBLE-ACUTE>;<CAPITAL>;IGNORE
<o">            <O//>;<DOUBLE-ACUTE>;<SMALL>;IGNORE
% <AA> is a separate letter in Danish
<AA>            <AA>;<NONE>;<CAPITAL>;IGNORE
<aa>            <AA>;<NONE>;<SMALL>;IGNORE
<A-A>           <AA>;<A-A>;<CAPITAL>;IGNORE
<A-a>           <AA>;<A-A>;<CAPITAL-SMALL>;IGNORE
<a-A>           <AA>;<A-A>;<SMALL-CAPITAL>;IGNORE
<a-a>           <AA>;<A-A>;<SMALL>;IGNORE
<AA'>           <AA>;<AA'>;<CAPITAL>;IGNORE
<aa'>           <AA>;<AA'>;<SMALL>;IGNORE
reorder-end
END LC_COLLATE

LC_CTYPE
copy "en_DK"
END LC_CTYPE

LC_MONETARY
int_curr_symbol         "<D><K><K><SP>"
currency_symbol         "<k><r>"
mon_decimal_point       "<,>"
mon_thousands_sep       "<.>"
mon_grouping            3;3
positive_sign           ""
negative_sign           "<->"
int_frac_digits         2
frac_digits             2
p_cs_precedes           1
p_sep_by_space          2
n_cs_precedes           1
n_sep_by_space          2
p_sign_posn             4
n_sign_posn             4
END LC_MONETARY

LC_NUMERIC
decimal_point           "<,>"
thousands_sep           "<.>"
grouping                3;3
END LC_NUMERIC

LC_TIME
abday       "<s><o//><n>";"<m><a><n>";/
            "<t><i><r>";"<o><n><s>";/
            "<t><o><r>";"<f><r><e>";/
            "<l><o//><r>"
day         "<s><o//><n><d><a><g>";/
            "<m><a><n><d><a><g>";/
            "<t><i><r><s><d><a><g>";/
            "<o><n><s><d><a><g>";/
            "<t><o><r><s><d><a><g>";/
            "<f><r><e><d><a><g>";/
            "<l><o//><r><d><a><g>"
abmon       "<j><a><n>";"<f><e><b>";/
            "<m><a><r>";"<a><p><r>";/
            "<m><a><j>";"<j><u><n>";/
            "<j><u><l>";"<a><u><g>";/
            "<s><e><p>";"<o><k><t>";/
            "<n><o><v>";"<d><e><c>"
mon         "<j><a><n><u><a><r>";/
            "<f><e><b><r><u><a><r>";/
            "<m><a><r><t><s>";/
            "<a><p><r><i><l>";/
            "<m><a><j>";/
```

```
            "<j><u><n><i>";/
            "<j><u><l><i>";/
            "<a><u><g><u><s><t>";/
            "<s><e><p><t><e><m><b><e><r>";/
            "<o><k><t><o><b><e><r>";/
            "<n><o><v><e><m><b><e><r>";/
            "<d><e><c><e><m><b><e><r>"
d_t_fmt     "<%><a><SP><%><d><SP><%><b><SP><%><Y><SP><%><T><SP><%><Z>"
d_fmt       "<%><d><-><%><m><-><%><Y>"
t_fmt       "<%><T>"
am_pm       "";""
t_fmt_ampm  ""
END LC_TIME

LC_MESSAGES
yesexpr     "<<(><1><J><j><Y><y><)/>><.><*>"
noexpr      "<<(><0><N><n><)/>><.><*>"
END LC_MESSAGES
```

E.6.2 Irish example

```
% Irish Gaelic language locale for Ireland
% Source: NSAI
% Address: Everson Gunn Teoranta, 15 Port Chaeimhghein <I'>ochtarach,
%          Baile <A'>tha Cliath 2, <E'>IRE
% Contact: Michael Everson
% Email: everson@indigo.ie
% Tel: +353 1 478-2597
% Fax: +353 1 478-2597
% Language: ga
% Territory: IE
% Revision: 0.5
% Date: 1996-01-28
% Users: general
% Repertoiremap: mnemonic,ds
% Charset: ISO_8859-1:1987
% Distribution and use is free, also for commercial purposes.

% Clause 1, Alphanumeric deterministic ordering
% Cuid 1, Rang<u'> docht alfa-uimhri<u'>il
LC_COLLATE
% Based on European multilingual ordering (to replace "en_DK")
% Accented characters are ordered as their unaccented counterparts.
% In the case of equality after this comparison has been made,
% unaccented characters are ordered before accented characters,
% and capital letters are ordered before small letters.
%
% Bunaithe ar an ord iltheangach Eorpach (in <a'>it "en_DK")
% D<e'>antar litreacha ar a bhfuil comhartha<i'> idirdhealaitheacha a
% rang<u'> d<i'>reach ar n<o'>s na coda eile. Tugtar t<u'>s <a'>ite don
% litir lom laistigh den rang<u'> sin, agus tugtar t<u'>s <a'>ite
% laistigh den rang<u'> sin do litreacha m<o'>ra thar litreacha beaga.
copy "en_DK"
END LC_COLLATE


% Clause 2, Classification of characters
% Cuid 2, Aicm<i'> carachtar
LC_CTYPE
% Based on European multilingual classification (to replace "en_DK")
% Bunaithe ar na haicm<i'> carachtar iltheangacha Eorpacha
% (in <a'>it "en_DK")
copy "en_DK"
END LC_CTYPE


% Clause 3, Numeric formatting
% Cuid 3, Form<a'>id<i'> uimhri<u'>la
LC_NUMERIC
decimal_point           "<.>"
thousands_sep           "<,>"
grouping                3;3
END LC_NUMERIC


% Clause 4, Monetary formatting
% Cuid 4, Form<a'>id<i'> airgid
LC_MONETARY
% Currency symbols (int_curr_symbol < ISO 4217)
% Comhartha<i'> airgeadais (int_curr_symbol < ISO 4217)
int_curr_symbol         "<I><E><P><SP>"
currency_symbol         "<Pd>"
% Normal numeric writing
% Gn<a'>thscr<i'>bhinn uimhri<u'>il
mon_decimal_point       "<.>"
mon_thousands_sep       "<,>"
mon_grouping            3;3
% Plus and minus symbols
% Comhartha<i'> m<o'>ide agus l<u'>ide
positive_sign           ""
```