

DRAFT - An alternative to the current ISO/IEC 9995-3 (*Work in progress*)

Draft Version 3 (revised) – 2008-03-02

Karl Pentzlin (karl.pentzlin@europatastatur.de)

0. Preface (*not to be a part of the final text*)

The current version of ISO/IEC 9995-3 intends to enable the input of a character repertoire as defined by collection 281 (MES-1) specified in amendment 1 to ISO/IEC 10646-1:2000.

Today, this character collection is not suited as base for a standardized keyboard layout, for the following reasons:

- The collection 281 is based of the ISO/IEC 6937, which was developed in the 1970s for "telematic services", i.e. for communication purposes like the long forgotten Telex successor "Teletex". It was not its primary goal to act as a well thought set for an international keyboard.
- ISO/IEC 6937 does not adhere to the same encoding principles as Unicode, which is prevalent in today's data processing systems. Especially, it is missing the mechanism of Unicode's combining characters. The ISO/IEC 6937 conforming mechanism of forming an accented letter by base letter + backspace + spacing accent is no longer possible with Unicode.
- Moreover, some characters of the collection 281 are obsolete legacy today which should not burden an actual keyboard design.
- The last 30 years yielded the need for some more characters (e.g., the introduction of the Latin alphabet in Azerbaijan revived the Jaňalif character Ə/ə).
- Additionally, the collection 281 is defective (e.g., it contains the characters Đ/đ, Ʀ/Ƨ and Ɖ/đ for Sámi, but not Ğ/g, Ʒ/ʒ and Ǟ/ǟ).
- As the name "MES-1" ("Multilingual European Subset 1") suggests, the larger part of the world is not considered (especially Vietnamese, but also most "minority languages" even if they write Latin). *When Sámi is taken into account in an international standard, why not Yorùbá or Comanche?*

0a. The character repertoire (*not to be a part of the final text*)

The character repertoire as specified implicitly by this document (consisting of all characters listed as associated with any key) is designed to met the following main requirements:

- a. All current languages which use the Latin script should be covered.
- b. To enable writing of proper names (e.g. in reference lists) and geographical names correctly, all transliteration systems for major current non-Latin languages into Latin should be covered.
- c. All symbols and punctuation marks which occur in good typography should be covered. *This includes ZWNJ, e.g. to prevent the f-l ligature in German »Schilfinsel« according to the orthographic rules, unlike the Soft Hyphen, which must not prevent a f-f ligature in »Affe« when applied within the »ff«.*
- d. All symbols which occur in business correspondence should be covered.

Additionally, it meets the following:

- e. *It contains the few letters and symbols (long s, long r, Tironian et) needed for the script variants Gaelic and Fraktur, which despite to their historical appeal have some contemporary use.*
- f. *It contains a small selection of historic letters (e.g. for Old English) and transliteration letters for historic scripts (for Egyptian hieroglyphs and Gothic), as these may be used in popular texts and texts for school use.*
- g. *It contains some characters for compatibility reasons.*

Not explicitly covered are special large sets like:

- a. phonetic characters like IPA
- b. mathematics

0b. The design of an international keyboard extension

(not to be a part of the final text)

The goal of ISO/IEC 9995-3 is to provide a possibility to type the additional character repertoire using any keyboard which adheres to some prerequisites, without referring to the actual layout.

Especially, it is required that there are the Latin letters (either as primary or as a secondary group), together with some other universal characters (like digits).

Rather than relying on physical positions, this proposal relies to the positions which the specific characters have on the basic layout. It seems far easier to communicate "to type æ, type AltGr+a" regardless whether the basic layout is QWERTY or AZERTY, rather than "to type æ, type AltGr together with the second key in the third row".

A quick survey of some Latin keyboard layouts which adhere to different national standards or conventions show that there are always different keys for the letters A...Z, the digits 0...9 and the punctuation marks dot and dash (while there are layouts which have the comma on the shifted position of the dot key). Therefore, these 38 characters, together with the ubiquitous function keys for Space, Tab and Backspace, are selected as base reference for the keyboard associations specified here (see Clause 4).

To provide a complete set of symbols together with a complete set of letters, two groups A (for additional letters and diacritical marks) and B (for additional punctuation marks and symbols) are provided. (For the group naming by letters, see Clause 4).

There are a set of diacritical marks (like acute accent or cedilla), together with some keys which appear as diacritical mark keys to the user, but are in fact additional group selectors. These are the diagonal stroke, the horizontal stroke, the hook above and the hook below.

As letters with these marks are encoded in Unicode only as composed forms (unless letters with true diacritics which are representable in Unicode as sequences of separately encoded base letter + diacritic), those characters are supplied as their own groups.

Likewise, the superscripting and subscripting of characters is supplied like "diacritic functions" for the user which are in fact separate groups.

Thus, in Clause 4 there is a total of 8 groups specified.

Most existing standards require the diacritical marks to be entered before the base character ("dead keys"). This principle is fully retained in this proposal even for sequences of more than one diacritic; it is specified how such input is to be converted into appropriate Unicode sequences (see Clause 5).

Additionally, function keys are specified to enter any valid Unicode character (see Clause 12), to provide a standard way for this rather than relying unstandardized special functions of operating systems or any text processing software.

On a standard US English keyboard, Group A may be selected by pressing the Right Alt (AltGr) key together with the key associated to the wanted character (and together with one of the Shift keys when selecting a character from Group A Level 2).

The combination Right Alt + Comma may select Group B for the next key (or combination of a character key + a Shift key) pressed after releasing that combination.

Thus, the selection mechanism for Group B also acts like a "diacritic function", like all groups except Group A.

0c. Superfluous and obsolete characters in collection 281

(not to be a part of the final text)

In the current version of Part 2 of ISO 9995 is stated:

For the input of graphic character repertoire of collection 281 (titled MES-1) as specified in amendment 1 to ISO/IEC 10646:1-2000, a Common Secondary Group Layout (to be used as group 2) is specified in ISO/IEC 9995-3.

The collection 281 is:

U+00..: 20-7E A0-FF

U+01..: 00-13 16-2B 2E-4D 50-7E

U+02..: C7 D8-DB DD

U+20...: 15 18-19 1C-1D AC

U+21...: 22 26 5B-5E 90-93

U+26...: 6A

This collection is no longer considered as a base for a character repertoire for an international keyboard extension (see clause 0). Especially, the following characters appear to be superfluous (*and are contained in the character repertoire specified by this document, if any, only for compatibility reasons*):

U+00A6 BROKEN BAR:

No real use attested beyond special mathematical-logical applications. Historic variant of U+007C VERTICAL LINE.

U+00AC NOT SIGN

Mathematical symbol without any attested business use. Not to be included into a repertoire which does not cover an appropriate large set of symbols for mathematics or formal logic.

U+0132 LATIN CAPITAL LIGATURE IJ

U+0133 LATIIN SMALL LIGATURE IJ

These are nowadays written as separate letters (see the detailed discussions in the Unicode mailing list archive).

U+0138 LATIN SMALL LETTER KRA

Was used in a former Greenlandic orthography, now obsolete there.

U+013F LATIN CAPITAL LETTER L WITH MIDDLE DOT

U+0140 LATIN SMALL LETTER L WITH MIDDLE DOT

Included in Unicode only as legacy compatibility characters. The preferred representations for Catalan by Unicode are U+004C U+00B7 resp. U+006C U+00B7.

U+0149 LATIN SMALL LETTER N PRECEDED BY APOSTROPHE

Included in Unicode only as legacy compatibility character.

The appropriate Unicode representation of the Afrikaans letter 'n' is U+02BC U+006E.

U+2126 OHM SIGN

This Unicode character has a canonical equivalence to U+03A9 GREEK CAPITAL LETTER OMEGA. Therefore according to the Unicode rules, the latter code is to be preferred for the Ohm sign.

(Note: To the U+00B5 MICRO SIGN, such an argument does not apply, as it has only a compatibility equivalence to U+03BC GREEK SMALL LETTER MU.)

U+266A EIGHTH NOTE

No special use is attested for this symbol in plain text. Not to be included into a repertoire which does not cover an appropriate large set of iconic symbols.

U+00...: C0-C5 C7-CF D1-D6 D9-DD E0-E5 E7-EF F1-F6 F9-FD FF

U+01...: 00-0F 12-13 16-25 28-2B 2E-2F 34-37 39-3E 43-19 4C-4D 50-51 54-65 68-7E

These are precomposed letters, not needing to be enumerated as it is sufficient to have the constituent characters to be included in the repertoire

All precomposed letters are implicitly contained in a set which includes all characters generated by application of the Unicode Normalization Form NFC on any sequence of characters specified in the original repertoire.

0d. Changes over version 2 (2008-02-19) of this draft *(not to be a part of the final text)*

- Some clarifications added
- A list of possible substitutions added to the description of Group B (Clause 6)
- special accents moved from Group H to Group E (for easier typing).
- Minor addition to Group B:
Some typographical alternatives for common mathematical symbols contained in ASCII (which, as universal symbols, commonly undergo some design compromises in fonts).

1. Scope

Within the general scope described in part 1 of ISO/IEC 9995, this DRAFT defines the allocation on a keyboard of a set of graphic characters which, when used in combination with an existing national version keyboard layout, allows the input of a minimum character repertoire as defined below.

This repertoire is intended to contain all characters needed to write all contemporary languages using the Latin script, together with standardized Latin transliterations of some major languages using other scripts. Also, it contains all symbols and punctuation marks contained in ISO 8859-1, together with some selected other ones commonly used in typography and office use.

This DRAFT is primarily intended for word-processing and text-processing applications.

2. Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of this part of ISO/IEC 9995. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this part of ISO/IEC 9995 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/IEC 646:1991, Information technology — ISO 7-bit coded character set for information interchange.
- ISO/IEC 9995-1:1994, Information technology — Keyboard layouts for text and office systems — Part 1: General principles governing keyboard layouts.
- ISO/IEC 10646-1: 2003 Information technology – Universal Multiple-Octet Coded Character Set (UCS) – Part 1: Architecture and Basic Multilingual Plane.
- Unicode 5.1 ... *exact reference TBD*
- Transliteration standards: ... *TBD, tentative and incomplete list:*
 - ISO 9 — Cyrillic*
 - ISO 233, DIN 31635 — Arabic*
 - ISO 259 — Hebrew*
 - ISO 843 — Greek*
 - ISO 3602 — Japanese*
 - ISO 7098 — Chinese*
 - ISO 9984 — Georgian*
 - ISO 9985 — Armenian*
 - ISO 11940 — Thai*
 - ISO 11941 — Korean*
 - ISO 15919 — Indic scripts*

3. Terms and Definitions

"actuate" a character: selecting a character by selecting the appropriate group and level (if necessary) and pressing the key itself.

"associated with": A key is associated with a character if it is used to enter that character, regardless of any level or group selection to be done before.

"base character": any graphic symbol which is not a diacritical mark and not a diacritical-neutral character.

"diacritical key": key associated with a diacritical mark (see Clause 5), when actuating this diacritical mark.

"diacritical-neutral character": any Unicode character which may influence the appearance of other characters without having any graphic representation itself. Contained in the supplementary character collection, this are U+200C ZERO WIDTH NON-JOINER (ZWNJ) and U+034F COMBINING GRAPHEME JOINER (CGJ). Other examples are U+200D ZERO WIDTH JOINER (ZWJ) or any Unicode variant selectors.

"digit key": a key "associated with" any digit 0 ... 9.

"non-diacritical key": key associated with a graphic symbol which is not a diacritical mark and not a diacritical-neutral character, when actuating this graphic symbol.

"supplementary groups": The groups defined in this document.

"supplementary character collection": All characters contained in any of the supplementary groups.

"symbol" (if not used within the term "graphic symbol" as defined in ISO/IEC 9995-1): Any graphic symbol which is neither a letter nor a punctuation mark.

Additionally, for the purposes of this DRAFT, the terms and definitions given in ISO/IEC 9995-1 apply

4. Conformance

The layout of a keyboard conforms to this DRAFT if it meets all of the following conditions:

- It is intended to output valid Unicode characters and valid sequences thereof.
- It has at least 41 distinct keys which are associated with:
 - the Latin uppercase letters A ... Z: U+0041 ... U+005A,
 - the digits 0 ... 9: U+0030 ... U+0039,
 - the full stop ("dot"): U+002E,
 - the hyphen-minus ("dash"): U+002D
 - the space: U+0020,
 - the Backspace function,
 - the Tab function.
- The following characters can be entered without recurring to the means specified in this DRAFT. U+0020 ... U+0022, U+0025 ... U+003F, U+0041 ... U+005A, U+005F, U+0061 ... U+007A, i.e.:
Space ! " % & ' () * + , - . / 0...9 : ; < = > ? A...Z _ a...z
(this are the characters contained in all national variants of ISO 646).
- The following groups can be selected:
 - Group A (Additional characters and group selectors), with two levels
 - Group B (Additional symbols), with three levels (the third one optional).

(The groups are denoted by letters here, as the actual group numbers may depend on the actual keyboard design.)

Any statement of conformance to this International Standard shall be taken to imply that the complete character repertoire specified here has been implemented, unless a subset is explicitly declared, provided that all other requirements specified here are respected.

5. Diacritical marks

Diacritical marks are the characters contained in the supplementary character collection specified here which are combining characters as defined by Unicode. Also, any character in a Private Use Area of Unicode may be treated as a diacritical marks depending of the operating system.

Diacritical marks appear above or below certain letters, and all of them are non-spacing characters.

- Actuating a diacritical mark or a sequence starting with a diacritical mark followed by any diacritical marks and/or diacritical-neutral characters, followed by actuating a base character key or any function key which is not a group or level selector, shall generate a sequence of Unicode character as follows:
 1. A character sequence is temporarily generated consisting of the actuated base character first (or, if a function key which is not a group or level selector was operated last, a U+00A0 NON-BREAKING SPACE instead), followed by the diacritical marks and diacritical-neutral characters in the order as actuated;

2. if the base character entered was Space, and if the last actuated character before is a diacritical mark which has a spacing clone in Unicode, the Space is replaced by that spacing clone, while the character last actuated before is removed from the temporary sequence,
3. then, on the temporary sequence, the Unicode NFC form is applied,
4. then, the character sequence thus generated is output,
5. then, if the last operated key was a function key which is not a group or level selector, that key will be treated accordingly.

It is recommended that the method used for the deletion of a character should also be used to cancel a partially constructed character, such as a diacritical mark without a following letter or a following Space character.

6. Layout Principles

- Rather than recurring to absolute positions on the keyboard, the additional characters are assigned to the 41 keys mentioned in Clause 4 which are denoted by the associated character enclosed in brackets, namely to [A] ... [Z], [0] ... [9], [.] , [-], [Space] and [BS] (the latter denoting the key associated with the Backspace function).

This implies that this DRAFT defines a means to identify the keys needed for the additional characters, rather than to define absolute locations.

- The character repertoire is organized into 8 groups (each having two levels, namely unshifted/shifted):

1. Group A: Additional characters and group selectors
2. Group B: Additional punctuation marks and symbols
3. Group C: Superscript characters
4. Group D: Subscript characters
5. Group E: Characters with diagonal stroke and special accents
6. Group F: Characters with horizontal stroke
7. Group G: Characters with hook above and other special characters
8. Group H: Characters with hook below and other special characters

Groups A and B are to be selected by means of the base keyboard layout. The selection keys for the other groups are contained in group A, in a manner that pressing and subsequently releasing these group selectors selects the effect of the next key pressing.

- Diacritical marks to be applied above the base letters are associated to level 1 (unshifted) positions (as these are the most frequent ones); such marks applied below the base letters are associated to level 2 (shifted) positions.
(This also corresponds to the fact that the low line U+005F is found on a shifted position on some common keyboard layouts.)
- Accordingly, the Group C/G selectors ("something above") are found on unshifted positions, whereas the Group D/H selectors ("something below") are found on shifted positions.
- The diacritical marks resembling dot and dash are associated with [.] and [-], respectively, in Group A.

All other diacritical marks are associated with number keys (instead of lumping all diacritical marks on a small group of keys). There are "accents" in Group A and "special accents" in Group E (the latter for special languages, transliteration, and other special purposes).

Thus, diacritical marks may be easily referenced to like "high/low [special] accent no. xxx" (besides "high/low dot/dash accent") without having to remember the real names (macron, ogonek, cedilla, etc.) or the design details.

- The fact that only a limited character set is required for the base layout (see Clause 4) may lead to a certain duplication of graphic characters between the base layouts and the layout of the additional

groups specified here. However, it allows the graphic characters of the groups specified here and their allocation to keys to be always the same for their use with any established Latin group layout.

6a. Character notation in the Group tables

In the Group tables, the notation of every character consists of:

- the character itself
(*TBD: display each character using a font which contains it*),
- the Unicode value,
- a parenthesized letter denoting the reason for including the character into the repertoire,
- following in some cases by the complete or shortened Unicode name.
- For some characters, there is provided (in parentheses) one example of the languages which use it (such language examples are not intended to denote the only or most prominent of such languages) or another explanation of the use of that character.

This reason codes are:

- c letter or diacritical mark used in any current language.
- t letter or diacritical mark used in any transliteration of a current language.
- h letter or diacritical mark used in any historical language or a transliteration for a historical language.
- v letter, diacritical mark or symbol needed for a variant of the Latin script (Fraktur, Gaelic)
- s punctuation mark or symbol in common use (for mathematical symbols: in common use outside of mathematical texts).
- y Needed for correct typography.
- m Only included for compatibility reasons.
- x Included only because the character is in Unicode and meets the specification of the Group (e.g., Group C Level 1 contains "raised" characters like ², ³, ⁿ, thus all "raised" characters contained in Unicode and "naturally" associated with a letter key are included).
- o Other reason (see explanation in parentheses or before the table).

If more than one reasons apply, only the first one is shown according to the list above (unless the other one is of significant importance).

The reason code is marked with one or two asterisks on the following conditions:

- ** Character contained in ASCII.
- * Character contained in collection 281 (the base for the pre-2008 editions of ISO/IEC 9995-3) but not in ASCII.

Note: All characters which are *not* marked with a * or ** are missing in the pre-2008 editions of ISO/IEC 9995-3.

7. Group A: Additional characters and group selectors

A notation like [Group X] means: Group selector for Group X.

The functions [UnicodeDec] and [UnicodeHex] are defined in Clause 12.

Annotations in italics are examples of use; in no case exhaustively.

Key	Level 1 (unshifted)	Level 2 (shifted)
[1]	´ U+0301 (c) acute accent	, U+0326 (c) comma below (<i>Romanian</i>)
[2]	¨ U+0308 (c) diaeresis	¨ U+0324 (c) diaeresis below (<i>Hakka</i>)
[3]	` U+0300 (c) grave accent	ˆ U+0328 (c) ogonek (<i>Polish</i>)
[4]	~ U+0303 (c) tilde	˘ U+0330 (c (?), t) tilde below

[5]	ˆ U+030B (c) double acute accent	¸ U+0329 (c) vertical line below (<i>Yorùbá</i>)
[6]	˘ U+0306 (c) breve	¸ U+032E (t) breve below (<i>Arabic transliteration</i>)
[7]	ˇ U+030C (c) caron	_ U+0332 (c) low line (<i>Comanche</i>)
[8]	ˆ U+0302 (c) circumflex accent	¸ U+032D (c) circumflex accent below (<i>Venda</i>)
[9]	ˆ U+0309 (c) hook above (hoi)	¸ U+0327 (c) cedilla (<i>Turkish</i>)
[0]	° U+030A (c) ring above	¸ U+0325 (c (?); t) ring below
[.]	˙ U+0307 (c) dot above	¸ U+0323 (c) dot below (<i>Africa</i>)
[-]	˘ U+0304 (c) macron	_ U+0331 (c) macron below (<i>Africa</i>)
[Q]	[Group G]	[Group H]
[W]	ə U+01DD (c) (<i>Nigeria</i>)	Ɔ U+018E (c) (<i>Nigeria</i>)
[E]	ə U+0259 (c) (<i>Azerbaijani</i>)	Ɔ U+018F (c) (<i>Azerbaijani</i>)
[R]	ṙ U+027C (v) (<i>Gaelic</i>)	Ɔ U+0295 (c (?); t)
[T]	þ U+00FE *(c) (<i>Icelandic</i>)	Ɔ U+00DE *(c) (<i>Icelandic</i>)
[Y]	ß U+00DF *(c) (<i>German</i>)	U+1E9E (c) (<i>German special applications</i>)
[U]	8 U+0223 (c) (<i>Algonquin</i>)	8 U+0222 (c) (<i>Algonquin</i>)
[I]	ı U+0131 *(c) (<i>Turkish</i>)	İ U+0130 *(c) (<i>Turkish</i>)
[O]	œ U+0153 *(c) (<i>French</i>)	Œ U+0152 *(c) (<i>French</i>)
[P]	ᑭ U+0242 (c) (<i>Chipewyan</i>)	ᑭ U+0241 (c) (<i>Chipewyan</i>)
[A]	æ U+00E6 *(c) (<i>Danish</i>)	Æ U+00C6 *(c) (<i>Danish</i>)
[S]	ſ U+017F (v) (<i>Gaelic, Fraktur</i>)	ᑭ U+0294 (c) (<i>Nootka</i>)
[D]	ð U+00F0 *(c) (<i>Icelandic</i>)	Ð U+00D0 *(c) (<i>Icelandic</i>)
[F]	ʌ U+028C (c) (<i>Temne</i>)	ʌ U+0245 (c) (<i>Temne</i>)
[G]	ɑ U+0251 (c) (<i>Fe'fe'</i>)	U+2C6D (c) (<i>Fe'fe'</i>)
[H]	ε U+025B (c) (<i>Nigeria</i>)	Ɔ U+0190 (c) (<i>Nigeria</i>)
[J]	ɔ U+0254 (c) (<i>Nigeria</i>)	Ɔ U+0186 (c) (<i>Nigeria</i>)
[K]	ɔ U+028A (c) (<i>Africa</i>)	Ɔ U+01B1 (c) (<i>Africa</i>)
[L]	U+A78C (c) small Saltillo (<i>Mex.</i>)	U+A78B (c) capital Saltillo (<i>Mexico</i>)
[Z]	[Group C]	[Group D]
[X]	[Group E]	[Group F]
[C]	ɜ U+0292 (c) (<i>Sámi</i>)	ɜ U+01B7 (c) (<i>Sámi</i>)
[V]	ɛ U+01B9 (c) (<i>Africa</i>)	ɛ U+01B8 (c) (<i>Africa</i>)
[B]	η U+019E (c) (<i>Lakota</i>)	ŋ U+0220 (c) (<i>Lakota</i>)
[N]	η U+014B *(c) (<i>Sámi</i>)	Ŋ U+014A *(c) (<i>Sámi</i>)
[M]	ɿ U+0283 (c) (<i>Africa</i>)	Σ U+01A9 (c) (<i>Africa</i>)
[Space]	[NBSP] U+00A0 *(s)	[NNBSP] U+202F (s)
[BS]	[ZWNJ] U+200C (y/v) (<i>German</i>)	[CGJ] U+034F (o) (<i>special lexical marking</i>)
[TAB]	[UnicodeDec]	[UnicodeHex]

On keyboard layouts where the [I] key is also associated with the lowercase dotless I U+0131, and where is a separate key associated with the uppercase dotted I U+0130 and the lowercase i U+0069, there the Group A association specified here does not need to be applied.

For the Vietnamese horn, see Group G applied to o/O and u/U.

Table of the spacing clones of the diacritics associated with the digit keys and [., [-] (which are yielded by entering the diacritic + SPACE):

Key	Level 1 (unshifted)	Level 2 (shifted)
[1] + [Space]	´ U+00B4 *(o) acute accent	, U+002C *(s) comma
[2] + [Space]	¨ U+00A8 *(o) diaeresis	-- (<i>diaeresis below</i>)
[3] + [Space]	` U+0060 *(o) grave accent	U+02DB *(o) ogonek
[4] + [Space]	U+02DC (o) small tilde (<i>not U+007E</i>)	-- (<i>tilde below</i>)
[5] + [Space]	U+02DD *(o) double acute accent	U+02CC (o) low vertical line
[6] + [Space]	U+02D8 *(o) breve	-- (<i>breve below</i>)
[7] + [Space]	U+02C7 *(o) caron	-- (<i>low line</i>)
[8] + [Space]	U+02C6 (o) circumflex (<i>not U+005E</i>)	U+A788 (o) low circumflex accent
[9] + [Space]	-- (<i>hook above</i>)	U+00B8 *(o) cedilla
[0] + [Space]	U+02DA *(o) ring above	U+02F3 (o) low ring
[.] + [Space]	U+02D9 *(o) dot above	U+002E *(s) full stop
[-] + [Space]	U+00AF *(o) macron	U+02CD (o) low macron

8. Group B: Additional punctuation marks and symbols

This group contains a level 3. This level contains:

- fractions beyond $\frac{1}{4}, \frac{1}{2}, \frac{3}{4}$
- some typographical alternatives for common mathematical symbols contained in ASCII (which, as universal symbols, commonly undergo some design compromises in fonts).
- currency symbols (except those which are contained in ISO 8859-1; these are contained in Level 1),
- some other symbols which are included into the supplementary character collection only because they are contained in ISO 8859-1 (namely, U+00A6, U+00AC, U+00AD).

If a keyboard design does not provide a Level 3 the characters listed here as Level 3 may be omitted.

Possible substitutions:

If the Group 1 of the keyboard already contains U+0024 \$, Group B Level 1 Key [S] may be assigned to U+00AD soft hyphen instead of U+0024 \$.

If the Group 1 of the keyboard already contains the brackets and braces (U+005B, U+005D, U+007B, U+007D), Group B level 1 Keys [D]/[F]/[G]/[H] may be assigned to $\frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}$ (+215B, U+215C, U+215D, U+215E) instead.

These substitutions especially may apply if the Level 3 of Group B is omitted.

Key	Level 1 (unshifted)	Level 2 (shifted)	Level 3
[1]	$\frac{1}{4}$ U+00BC *(s)	± U+00B1 *(s)	$\frac{1}{3}$ U+2153 (s)
[2]	$\frac{1}{2}$ U+00BD *(s)	≈ U+2248 (s)	$\frac{2}{3}$ U+2154 (s)
[3]	$\frac{3}{4}$ U+00BE *(s)	≠ U+2260 (s)	$\frac{3}{8}$ U+215C *(s)
[4]	ˆ U+2036 (s) rev. db. prime	˘ U+2035 (s) reverse prime	˜ U+223C (y) tilde operator

[5]	" U+2033 (s)	' U+2032 (s)	ⅆ U+215D *(s)
[6]	„ U+201E (s)	, U+201A (s)	– U+2212 (y) minus sign
[7]	“ U+201C *(s)	‘ U+2018 *(s)	⅔ U+215E *(s)
[8]	” U+201D *(s)	’ U+2019 *(s)	⅓ U+215B *(s)
[9]	« U+00AB *(s)	‹ U+2039 (s)	/U+2215 (y) division slash
[0]	» U+00BB *(s)	› U+203A (s)	/U+2044 (y) fraction slash
[.]	· U+00B7 *(s)	@ U+0040 *(s)	* U+2217 (y) asterisk operator
[-]	– U+2013 (s) (<i>German</i>)	— U+2014 (s) (<i>USA</i>)	- U+00AD *(s) soft hyphen
[Q]	∅ U+2300 (s) diameter	— U+2015 *(s)	¬ U+00AC *(m)
[W]	† U+2020 (s)	‡ U+2021 (s)	⚡ U+20A9 (s)
[E]	€ U+20AC *(s)	⌘ U+00A4 *(s)	€ U+20A0 (s)
[R]	° U+00B0 *(s)	® U+00AE *(s)	Rs U+20A8 (s)
[T]	^ U+005E *(s)	™ U+2122 (s)	₹ U+20AE (s)
[Y]	₸ U+00A5 *(s)	… U+2026 (s)	Kazhak tenge (encoding in progress)
[U]	¿ U+00BF *(s)	‰ U+2030 (s)	
[I]	¡ U+00A1 *(s)	§ U+00A7 *(s)	
[O]	º U+00BA *(s)	Ω U+03A9 (s) <i>replaces *U+2126</i>	
[P]	¶ U+00B6 *(s)	© U+2117 (s)	₱ U+20B1 (s)
[A]	ª U+00AA *(s)	½ U+214D (s) (<i>Norwegian</i>)	₣ U+20B3 (s)
[S]	\$ U+0024 *(s)	™ U+2120 (s) service mark	₪ U+20AA (s)
[D]	{ U+007B *(s)	← U+2190 *(s)	¤ U+20AB (s)
[F]	} U+007D *(s)	→ U+2192 *(s)	₯ U+20A3 (s)
[G]	[U+005B *(s)	≤ U+2264 (s)	₧ U+20B2 (s)
[H]] U+005D *(s)	≥ U+2265 (s)	₹ U+20B4 (s)
[J]	(U+27E8 (s) left angle bracket	≪ U+226A (s) much less	₺ U+266A *(m)
[K]) U+27E9 (s) right angle brck.	≫ U+226B (s) much greater	₠ U+20AD (s)
[L]	£ U+00A3 *(s)	¢ U+00A2 *(s)	₡ U+20B5 (s)
[Z]	↗ U+2197 (s) NE arrow	↘ U+2198 (s) SE arrow	₢ U+20B0 (s)
[X]	× U+00D7 *(s)	÷ U+00F7 *(s)	₣ U+20A2 (s)
[C]	% U+2105 (s)	© U+00A9 *(s)	₤ U+20A1 (s)
[V]	~ U+007E *(s)	U+007C *(s)	₶ U+00A6 *(m)
[B]	♭ U+266D (s) musical flat sign	♯ U+266F (s) (<i>for music titles</i>)	₷ U+0E3F (s)
[N]	\ U+005C *(s)	№ U+2116 (s)	₸ U+20A6 (s)
[M]	µ U+00B5 *(s)	ḡ Î U+204A (v) Tironian et	₹ U+2133 (v) German Mark
[SP]	# U+0023 *(s)		

9. Groups C/D

Group C: Superscript characters

Group D: Subscript characters

"Level 1" means "Level 1 (unshifted)", "Level 2" means "Level 2 (shifted)".

Key	Group C Level 1	Group C Level 2	Group D Level 1	Group D Level 2
[1]	¹ U+00B9 *(s)		U+2081 (x)	
[2]	² U+00B2 *(s)		U+2082 (x)	
[3]	³ U+00B3 *(s)		U+2083 (x)	
[4]	U+2074 (s)		U+2084 (x)	
[5]	U+2075 (s)		U+2085 (x)	
[6]	U+2076 (s)		U+2086 (x)	
[7]	U+2077 (s)		U+2087 (x)	
[8]	U+2078 (s)		U+2088 (x)	
[9]	U+2079 (s)		U+2089 (x)	
[0]	U+2070 (x)		U+2080 (x)	
[.]				
[-]				
[Q]				
[W]	^w U+02B7 (x)	^W U+1D42 (x)		
[E]	^e U+1D49 (x)	^E U+1D31 (x)	^e U+2091 (x)	
[R]	^r U+02B3 (x)	^R U+1D3F (x)	^r U+1D63 (x)	
[T]	^t U+1D57 (x)	^T U+1D40 (x)		
[Y]	^y U+02B8 (x)			
[U]	^u U+1D58 (x)	^U U+1D41 (x)	^u U+1D64 (x)	
[I]	U+2071 (x)	^I U+1D35 (x)	ⁱ U+1D62 (x)	
[O]	^o U+1D52 (x)	^O U+1D3C (x)	^o U+2092 (x)	
[P]	^p U+1D56 (x)	^P U+1D3E (x)		
[A]	^a U+1D43 (x)	^A U+1D2C (x)	^a U+2090 (x)	
[S]	^s U+02E2 (x)			
[D]	^d U+1D48 (x)	^D U+1D30 (x)		
[F]	^f U+1DA0 (x)			
[G]	^g U+1D4D (x)	^G U+1D33 (x)		
[H]	^h U+02B0 (t)	^H U+1D34 (t)		
[J]	^j U+02B2 (x)	^J U+1D36 (x)	U+2C7C (x)	
[K]	^k U+1D4F (x)	^K U+1D37 (x)		
[L]	^l U+02E1 (x)	^L U+1D38 (x)		
[Z]	^z U+1DBB (x)			
[X]	^x U+02E3 (x)		^x U+2093 (x)	

[C]	ˆ U+1D9C (x)			
[V]	˘ U+1D5B (x)	U+2C7D (x)	˘ U+1D65 (x)	
[B]	ˆ U+1D47 (x)	ˆ U+1D2E (x)		
[N]	ˆ U+207F (c) (Minnan)	ˆ U+1D3A (x)		
[M]	ˆ U+1D50 (x)	ˆ U+1D39 (x)		
[SP]	ˆ U+2191 *(s)		˘ U+2193 *(s)	

9a. Groups E/F

Group E:

Characters with diagonal stroke (associated with letter keys)

Special accents (associated with digit keys and [.] / [-]): diacritical marks for Minnan and for special applications, which do not fit into group A).

Group F: Characters with horizontal stroke (associated with letter keys)

(does not contain any characters associated with digit keys)

"Level 1" means "Level 1 (unshifted)", "L.2" means "Level 2 (shifted)".

Remark: Sençoten uses only capital letters, therefore the appropriate small letters have the reason code (x) instead of (c).

Key	Group E: Level 1 (unshifted)	Group E: Level 2 (shifted)
[1]	‘ U+0312 (h) turned comma above	
[2]	’ U+0313 (c) comma above (Minnan)	
[3]	‘ U+0314 (h) reversed comma above	
[4]	’ U+031B (c) horn (Vietnam; Thai transliteration)	
[5]	ˆ U+030F (o) double grave (Croat poetry)	˘ U+0347 (t) equals sign below
[6]	ˆ U+0311 (o) inverted breve (Croat poetry)	
[7]	ˆ U+0361 (t) double inverted breve	˘ U+035C (t) double breve below
[8]	ˆ U+0360 (t) double tilde	˘ U+035F (t) double macron below
[9]	ˆ U+030D (c) vertical line above (Minnan)	
[0]	ˆ U+0310 (t) candrabindu	
	ˆ U+0358 (c) dot above right (Minnan)	

Table of the spacing clones of the diacritics associated with the digit keys and [.] , [-] (which are yielded by entering the diacritic + SPACE):

Key	Group E: Level 1 (unshifted)	Group F: Level 2 (shifted)
[1] + [Space]	U+02BB modif. turned comma	
[2] + [Space]	U+02BC modif. apostrophe	
[3] + [Space]	U+02BD modif. reversed comma	
[4] + [Space]	--	

[5] + [Space]	--	--
[6] + [Space]	--	
[7] + [Space]	U+2040 character tie	U+203F undertie
[8] + [Space]	U+2053 swung dash	--
[9] + [Space]	U+02C8 vertical line	
[0] + [Space]	--	
[.] + [Space]	--	
[-] + [Space]	--	

Key	Group E :Level 1	Group E: L.2	Group F: Level 1	Group F: L.2
[Q]				
[W]				
[E]	ø U+0247 (x)	Ǝ U+0246 (x)		
[R]			ƚ U+024D (x)	ƚ U+024C (x)
[T]	ʄ U+2C66 (x) (Sencóten)	Ƨ U+023E (c)	Ƨ U+0167 *(c) (Sámi)	Ƨ U+0166 *(c)
[Y]			Ƨ U+024F (x)	Ƨ U+024E (x)
[U]			Ƨ U+0289 (c) (Comanche)	Ƨ U+0244 (c)
[I]			Ƨ U+0268 (c) (Africa)	Ƨ U+0197 (c)
[O]	ø U+00F8 *(c) (Danish)	Ø U+00D8 *(c)	ø U+0275 (c) (Africa)	Θ U+019F (c)
[P]			Ƨ U+1D7D (x)	Ƨ U+2C63 (x)
[A]	ǰ U+2C65 (x) (Sencóten)	ǰ U+023A (c)		
[S]				
[D]			đ U+0111 *(c) (Croatian)	Đ U+0110 *(c)
[F]				
[G]			ǰ U+01E5 (c) (Sámi)	ǰ U+01E4 (c)
[H]			ħ U+0127 *(x) (Maltese)	ħ U+0126 *(c)
[J]			ǰ U+0249 (x)	ǰ U+0248 (x)
[K]				
[L]	ł U+0142 *(c) (Polish)	ł U+0141 *(c)	ł U+019A (x) (Sencóten)	ł U+023D (c)
[Z]			z U+01B6 (x)	z U+01B5 (x)
[X]				
[C]	ç U+023C (x) (Sencóten)	Ç U+023B (c)		
[V]				
[B]			Ƨ U+0180 (h) (Old Saxon)	Ƨ U+0243 (h)
[N]				
[M]	ɱ U+20A5 (x)			

[SP]	/U+0338 (o)		-U+0336 (o)	
------	-------------	--	-------------	--

10. Group G: Characters with hook above and other special characters

Associated with the keys [0] ... [9], [.] and [-], this group contains "spacing modifier letters" (accent-like characters which occupy their own width rather than to display above or below a base character).

Also, associated with the keys [1] ... [5], but on level 2 (shifted), this group contains the Khoisan click letters.

Key	Level 1 (unshifted)	Level 2 (shifted)
[1]	' U+02BC (c) (Afrikaans)	ǀ U+01C0 (c (?)) (Khoisan click letter)
[2]	˘ U+02BB (c) (Hawaiian 'okina)	ǁ U+01C1 (c (?)) (Khoisan click letter)
[3]	ǁ U+0295 (t)	ǂ U+01C2 (c (?)) (Khoisan click letter)
[4]	ǃ U+0294 (t)	ǃ U+01C3 (c (?)) (Khoisan click letter)
[5]	' U+02B9 (t) (Cyrillic)	Ǆ U+0298 (c (?)) (Khoisan click letter)
[6]	" U+02BA (t) (Cyrillic)	
[7]	ˆ U+02BF (t) (Arabian; Hebrew)	
[8]	˙ U+02BE (t) (Arabian; Hebrew)	
[9]	' U+02C8 (t)	
[0]	, U+02CC (t)	
[.]	U+A789 (c) colon	
[-]	U+A78A (c) short equals sign	Ⲛ U+2E17 (h) (Coptic transliteration)
[Q]		
[W]	U+2C73 (c) (Africa)	U+2C72 (c)
[E]	◌ U+1D4A (t)	
[R]	ᚱ U+0280 (h) (Old Norse)	ᚱ U+01A6 (h)
[T]	ṭ U+0288 (c) (Africa)	Ṭ U+01AE (c)
[Y]	Ƴ U+01B4 (c) (Africa)	Ƴ U+01B3 (c)
[U]	ư U+01B0 (c) (Vietnam)	Ư U+01AF (c)
[I]	ɪ U+026A (c) (Africa)	
[O]	ơ U+01A1 (c) (Vietnam)	Ơ U+01A0 (c)
[P]	ɓ U+01A5 (c) (Africa)	Ɔ U+01A4 (c)
[A]	U+A723 (h) (Egypt hieroglyph transliteration)	U+A722 (h)
[S]		
[D]	ɗ U+0257 (c) (Africa)	Ɖ U+018A (c)
[F]		
[G]	ɠ U+0260 (c) (Africa)	Ɠ U+0193 (c)
[H]		
[J]	U+A725 (h) (Egypt hieroglyph transliteration)	U+A724 (h)
[K]	ƙ U+0199 (c) (Africa)	Ƙ U+0198 (c)

[L]		
[Z]		
[X]		
[C]	Ɔ U+0188 (c) (Africa)	Ɔ U+0187 (c)
[V]	ɔ U+028B (c) (Africa)	U U+01B2 (c)
[B]	ɓ U+0253 (c) (Africa)	ʙ U+0181 (c)
[N]		
[M]		

11. Group H: Characters with hook below and other special characters

This group has no characters associated with the digit keys.

Key	Level 1 (unshifted)	Level 2 (shifted)
[Q]	q U+024B (c) (Africa)	Q U+024A (c)
[W]	ƿ U+01BF (h) (Old English)	Ƶ U+01F7 (h)
[E]		
[R]	ɾ U+027D (c) (Africa)	Ŕ U+2C64 (c)
[T]	ɽ U+0288 (c) (Africa)	Ť U+01AE (c)
[Y]		
[U]	ɯ U+026F (c (?)) (Africa)	Ʊ U+019C (c (?))
[I]	ɪ U+0269 (c) (Africa)	Ƭ U+0196 (c)
[O]		
[P]		
[A]		
[S]		
[D]	ɖ U+0256 (c) (Africa)	Ɖ U+0189 (c)
[F]	ƒ U+0192 (c; s) (Africa)	F U+0191 (c)
[G]		
[H]	ɦ U+0195 (h) (Gothic)	Hu U+01F6 (h)
[J]	ɟ U+021D (h) (Old English)	Ʒ U+021C (h)
[K]	κ U+0138 *(m/h) (former Greenlandic)	
[L]	ł U+026D (x)	
[Z]	z U+0225 (h) (Middle High German)	Z U+0224 (h)
[X]		
[C]		
[V]	ƴ U+0263 (c) (Africa)	ƶ U+0194 (c)
[B]		
[N]	ɲ U+0272 (c) (Africa)	Ɲ U+019D (c)

[M]	m U+0271 (c) (Africa)	U+2C6E (c)
-----	-----------------------	------------

12. Function keys for entering any valid Unicode character

In Group A (see Clause 7), there are two function key declared: [UnicodeDec] and [UnicodeHex]. These are to enter any valid Unicode character, by entering their code point values as decimal resp. hexadecimal number.

- [UnicodeDec] puts the keyboard into a special state "Unicode Decimal Input". All actuations of keys associated with decimal digits are temporarily stored into a sequence representing a decimal number. When any other key is pressed, then, if the decimal number contains at least one digit and represents a valid Unicode value, then the according character will be output. If not, then an U+FFFD OBJECT REPLACEMENT CHARACTER will be output, followed by the entered sequence of decimal digits. In any case, the temporary sequence will be cleared.

Then, if the other key pressed is not an Enter key, a Decimal Separator key or a Space key, the special state "Unicode Decimal Input" will be cleared, and the other key will be treated normally.

If the other key pressed is an Enter key, the special state "Unicode Decimal Input" will be cleared, but the Enter key itself will not be processed further.

If the other key pressed is a Decimal Separator or a Space key, the special state "Unicode Decimal Input" will not be cleared, and the Decimal Separator key resp. the Space key itself will not be processed further.

Thus, the user can enter any sequence of valid Unicode characters by entering their decimal code values, separated by Space or decimal separator, and terminated by Enter.

[UnicodeHex] works accordingly. Hexadecimal digits are all decimal digits and A...F and a...f, not differentiating between upper and lower case.

Valid Unicode characters must have hexadecimal values between 0 and 10FFFF. Also, their value must not be in the intervals D800...DFFF (Unicode surrogate points) and FDD0...FDEF (Unicode noncharacters), and their value modulo hexadecimal 10000 must not be FFFE or FFFF (values guaranteed not to be a Unicode character at all by Unicode).

The operating system may provide more restrictions, e.g. usage of a code position in a specific version of Unicode.