

# Specification for a new Finnish Keyboard Layout

2005-12-14

**Prepared by the Keyboard Working Group under guidance from the “Kotoistus” Steering Group and the National “Kotoistus” Working Group on Cultural Diversity Issues in ICT.**

As part of its responsibilities, the Kotoistus group also handles certain subject areas of formal IT standardization (in CEN/ISSS and ISO/IEC JTC1), such as encoded character sets, keyboards, and internationalization issues in general, on behalf of and as authorized by the national standards body, the Finnish Standards Association SFS.

\*

## Introduction

This specification describes the design principles for a new Finnish-Swedish keyboard layout. In addition, this version is geared towards a specific implementation by Microsoft in the forthcoming Windows Vista. This is the predecessor for a more general specification that will be prepared using a more generic, system independent terminology (more in line with that of ISO/IEC 9995). The future specification will specify the same functionality (repertoire), but it will allow for more freedom in choosing how to produce the functions, including the positioning of certain characters and functions on the keytops.

The specification is based on the feedback from the national public review on the “Kotoistus” web site and from the international IT industry (Adobe, Apple, and Microsoft, in particular). The end result has been further affected by a compromise solution that has been worked out with the relevant Swedish standards committee, to facilitate also in practice a choice between their new standard and ours, since they both work with the same engravings (defined as mandatory in the Swedish standard), in spite of their very different design principles.

## Highlights

The new general purpose keyboard layout extends the functionality of the presently common SWE/FIN keyboard layout considerably.

In spite of this, there are no disturbing changes to the users of the base functions of the regular (i.e., desktop) keyboards. In particular, the number and placement of keys is the same as in the full size qwerty keyboards currently in use and there is no change to the current allocation of keys.

The keyboard layout is designed to produce UCS/Unicode encoding. If a more limited encoding is required, the functionality of the keyboard will have to be restricted correspondingly.

The design is open-ended and based on intuitively recognizable positioning of additional letters. It can thus be learned with a minimum effort on an as-needed basis.

## User Communities

The proposed keyboard layout is intended for use by the speakers of the majority languages of Finland and Sweden.

## Basic Goals

The keyboard layout is designed to meet three important basic goals.

First, it provides for easy entering of text in the Fenno-Scandinavian majority languages, i.e. Finnish, Swedish, Norwegian, and Danish, using a familiar keyboard layout.

Second, it is designed to give an easy, intuitively recognizable way to enter text in all of the official Fenno-Scandinavian regional and minority languages, including those with the requirement for additional characters (i.e., for Finland, Sweden and Norway: Northern Sámi, Southern Sámi, Lule Sámi, Inari Sámi, Skolt Sámi, and Romany as spoken in Finland, and for Denmark: Faroese and Kalaallisut aka Greenlandic).

As a third objective, it allows for relatively easy entering of particularly the names (of persons, places or products) with many Latin characters used in other languages, such as the official languages of the European Union (though not, of course, including Greek) plus the remaining Nordic language, Icelandic.

The Latin characters in modern use of the MES-2 (Multilingual European Subset 2 of the UCS/Unicode) repertoire are the basis for this design. The needs of writing Vietnamese names have also been taken into account due to the size of that immigration population.

## The Means

To achieve the goals, room is needed for non-decomposable characters in intuitively recognizable positions. For this purpose, almost all characters that are decomposable in UCS/Unicode terms are to be produced using a systematic approach utilizing dead keys to add diacritical marks. The letters å, ä, and ö are however kept in their currently allocated positions, for obvious reasons.

It is also of particular importance to the end users that specific punctuation is provided to support the orthographies of the languages in question.

Presently, the basic ASCII quotation marks and hyphen-dash are often used by various applications (correctly or incorrectly) to produce localized punctuation. These characters are now augmented by new punctuation marks that produce the desired result independently of the application programs. With these additions, the keyboard supports much better the proper punctuation of particularly the Nordic languages and English or German.

In addition, the no-break space is provided as an easy to use tool for preventing unsuitable line breaks when texts are formatted.

Thus, there would be considerable justification to call the keyboard layout “Nordic International”.

## Compositions

The keyboard layout in current use has two specific (dead) keys for diacritics. The new layout preserves these keys and provides for several additional diacritics.

In addition to the original dead keys, several keys gain dead key functionality when used in conjunction with the AltGr key (and possibly a Shift key). For example, the apostrophe key acts as a dead key for the caron when used together with the AltGr key. In the sequel, the term “dead key” refers to both the two dead keys proper and the dead key functionality of other keys.

N.B. For compatibility with the mandatory engravings of the Swedish standard, the dot above diacritic is assigned to the AltGr + Shift position in B10 (hyphen key) in addition to its more natural placement in that position in B09 (dot key).

This design is most naturally implemented by defining that a dead key logically produces a UCS/Unicode combining diacritical mark after the character that is typed next. Thus, pressing the upper dead key (with the engravings ´ and `) and then a letter key produces that letter followed by the combining acute accent (0301). However, the keyboard driver is expected to convert this to a precomposed character, such as á (00E1) or é (00E9), for the defined repertoire. For other than the defined repertoire, a decomposed sequence would normally be produced, but it may also be precomposed, if such a character exists in UCS/Unicode. This implies no change to current functionality for ISO Latin 1 characters, but it extends them in a well-defined, universal way.

Ultimately, multiple, different diacritical marks need to be attached to the same letter. Thus, e.g., diaeresis + caron + u would result in ũ (01DA as a precomposed character, or the decomposed sequence). Such functionality is needed for correct writing of e.g. Lithuanian and Vietnamese. From a practical point of view, this implementation may be staged, since the most common and urgent need is to support the “more familiar strange characters”.

### The stroke modifier key

The composition principle is extended to some characters that are not composite in the UCS/Unicode sense but can intuitively be seen as consisting of a basic Latin letter and a stroke, such as đ. The stroke in such characters is not a diacritic, but in the proposed keyboard layout, a special modifier key is used to create a character with a stroke. Although the stroke is not a decomposable element of a character, the keyboard driver can handle it in a manner comparable to the use of dead keys for diacritical marks. This makes typing such letters easy and natural, as soon as the user has learned the simple principle.

This approach makes it possible to type several Sámi letters using the modifier key (with AltGr) and a basic letter key. The characters involved are đ - d with stroke (0111, in Northern, Inari and Skolt Sámi), g - g with stroke (01E5, in Skolt Sámi), and t - t with stroke (0167, in Northern Sámi). – The stroke is also used to create e.g. ł - l with stroke (0142, in Polish) or ħ - h with stroke (0127, in Maltese), and can be used to write ø - o with stroke (00F8, also available directly on the ö key).

Since the stroke is not a decomposable element, all combinations of the modifier key and a letter that produce a character can be specified separately. (The logically simplest approach would be to allow any Latin letter with the word “stroke“ in its UCS/Unicode name to be produced that way, but only

those in the defined repertoire need be produced.) When used in any other than a valid combination, the modifier key would have no effect or could produce some stroke-like character.

In addition to the primary position for the stroke modifier key, a fallback position has been provided for use on those keyboards (mostly in some laptops) that don't have the primary position available.

## The Changes

The specification changes the meaning of some key sequences, but this will probably not cause much difficulties. At present, pressing e.g. the acute dead key and the c key produces ´c, whereas in the new scheme, it produces ´c. The sequence ´c would need to be produced by pressing the space key between the acute key and the c key. This is how users generally behave anyway even now.

The additional letters and diacritical marks are shown on the layout.

## Open-endedness

The layout is open ended. Only a minimum repertoire is defined.

In line with the current UCS/Unicode practice, the keyboard driver does not check the applicability of letters with combining diacritical marks. Thus, the responsibility for checking the acceptability of any resulting character belongs initially to the user (by means of viewing the display of what has been input) and ultimately to the application (which may set its own limitations to the repertoire – as always, since e.g. names in a population registry may not contain otherwise perfectly valid characters such as digits).

For the sake of consistent application of the design principle, no other decomposable letters than å, ä and ö have their own key position.

In Appendix 1, the specific new function (when compared with the common keyboard layout in current use) of each key is discussed.

Appendix 2 shows the proposed keyboard layout.

Appendix 3 defines the repertoire required for precomposed encoding with dead letter key input.

## Appendix 1:

The functions of the various data keys are as follows:

<i>Pos</i>	<i>Base function</i>	<i>+ Shift</i>	<i>Alt Gr function</i>	<i>Alt Gr + Shift</i>	<i>Comment</i>
E00	§	½	Modifier: stroke		Added stroke modifier. If no E00, see C09!
E01	1	!		¡ (00A1)	Added inverted exclamation mark.
E02	2	"	@	” (201D)	Added quotation mark.
E03	3	#	£	» (00BB)	Added quotation mark.
E04	4	α	\$	« (00AB)	Added quotation mark.
E05	5	%	‰ (2030)	“ (201C)	Added per mille sign + quotation mark.
E06	6	&	, (201A)	„ (201E)	Added low quotation marks.
E07	7	/	{		
E08	8	(	[		
E09	9	)	]		
E10	0	=	}	° (00B0)	Added degree sign.
E11	+	?	\	¿ (00BF)	Added inverted question mark.
E12	Cd: ´ acute	Cd: ` grave	Cd: ¸ cedilla (0327)	Cd: ı ogonek (0328)	Added diacritics.
D01	q	Q			
D02	w	W			
D03	e	E	€		
D04	r	R			
D05	t	T	þ (00FE)	Þ (00DE)	Added letter thorn.
D06	y	Y			
D07	u	U			
D08	i	I	ı (0131)		Added letter dotless i.
D09	o	O	œ (0153)	Œ (0152)	Added oe ligature.
D10	p	P	Cd: ˆ horn (031B)	Cd: ˆ hook above (0309)	Added diacritics (esp. for Vietnamese).
D11	å	Å	Cd: ˆ double acute (030B)	Cd: ˆ ring above (030A)	Added diacritics.
D12	Cd: ˆ diaeresis	Cd: ˆ circumflex	Cd: ˆ tilde	Cd: ˆ macron (0304)	Added diacritic.
C01	a	A	ə (0259)	Ə (018F)	Added letter schwa.
C02	s	S	ß (00DF)		Added letter sharp s.
C03	d	D	ð (00F0)	Ð (00D0)	Added letter eth.

C04	f	F			
C05	g	G			
C06	h	H			
C07	j	J			
C08	k	K	к (0138)		Added letter kra.
C09	l	L	Modifier: stroke		Added fallback pos. for modifier (for use when E00 not available).
C10	ö	Ö	ø (00F8)	Ø (00D8)	Added letter o with stroke.
C11	ä	Ä	æ (00E6)	Æ (00C6)	Added letter ae.
C12	'	*	Cd: ˇ caron (030C)	Cd: ˇ breve (0306)	Added diacritics.
B00	<	>			
B01	z	Z	з (0292)	З (01B7)	Added letter ezh.
B02	x	X	× (00D7)	· (00B7)	Added multiplication sign and middle dot.
B03	c	C			
B04	v	V			
B05	b	B			
B06	n	N	ŋ (014B)	Ŋ (014A)	Added letter eng.
B07	m	M	μ (00B5)	— (2014)	Added micro sign and em dash.
B08	,	;	' (2019)	' (2018)	Added single quotation marks.
B09	.	:	Cd: ˙ dot below (0323)	Cd: ˙ dot above (0307)	Added diacritics.
B10	-	–	– (2013)	Cd: ˙ dot above (0307)	Added en dash + diacritic.
A	Space		NBSP (00A0)		Added no-break space.

N.B. The layout intentionally leaves many positions unassigned, even though rather natural allocations could be presented. The need for additional assignments varies by the user, by the nature of the text, and by the program. Thus, they are best left unassigned in the basic layout, so that there is more room for varying useful customizations.