



**ISO/IEC JTC 1/SC 2/WG 3
7-bit and 8-bit codes and their extension
SECRETARIAT : ELOT**

DOC TYPE :	Expert Contribution
TITLE :	Soft Hyphen and some other characters
SOURCE :	Kent Karlsson
PROJECT:	--
STATUS :	
ACTION ID :	FYI
DUE DATE :	
DISTRIBUTION :	P, O and L Members of ISO/IEC JTC 1/SC 2 WG Conveners, Secretariats ISO/IEC JTC 1 Secretariat ISO/IEC ITTF
NO OF PAGES :	3

Contact 1: Secretariat ISO/IEC JTC 1/SC 2/WG 3 ELOT Mrs K.Velli
Acharnon 313, 111 45 Kato Patissia, ATHENS – GREECE
Tel: +30 1 21 20 307 Fax : +30 1 22 86 219 E-mail : kkb@elot.gr

Contact 2 : Convenor ISO/IEC JTC 1/SC 2/WG 3 Mr E.Melagrakis
Acharnon 313, 111 45 Kato Patissia, ATHENS – GREECE
Tel: +30 1 21 20 301 Fax : +30 1 22 86 219 E-mail: eam@elot.gr

SOFT HYPHEN and some other characters

The text concerning SOFT HYPHEN in the ISO/IEC 8859 series (and in ISO/IEC 6937) is unclear, and has been misinterpreted as disallowing SOFT HYPHEN if not immediately followed by a LINE FEED and/or CARRIAGE RETURN). See e.g.: <http://www.hut.fi/~jkorpe/shy.html>. This misinterpretation has been circulated as “the correct one” on one of the Linux mailing lists.

ISO/IEC 8859 says on SOFT HYPHEN: “A graphic character that is imaged by a graphic symbol identical with, or similar to, that representing hyphen, for use when a line break has been established within a word.”

The intent here is that the graphic symbol is to be used when “a line break has been established within a word, and that otherwise no graphic symbol is to be used.

The misinterpretation circulated is that the SOFT HYPHEN character itself should only be used if “a line break has been established within a word” and that line-break is then explicitly represented as a line feed (or carriage return). The text in the 8859 series on SOFT HYPHEN is unclear, and allows for this unintended interpretation. To avoid continued misinterpretation, the text on SOFT HYPHEN should be made clearer.

ISO/IEC 10646-1:2000 (annex H) contains no text at all on SOFT HYPHEN, nor on SPACE (8859 does), NO-BREAK SPACE (but on NARROW NO-BREAK SPACE), nor MONGOLIAN TODO SOFT HYPHEN. Annex F cannot reasonably be extended to cover all characters needing “special handling”, but some are covered by 8859 in an unclear fashion. Below are new texts for some special characters not already covered by annex H, or are unclear in the context of 8859. These texts are suggested also for the 8859 series, as well as 6937 (for those of these characters that are included there).

Suggested new texts:

SPACE (0020): SPACE (SP) is a graphic character that has a visual representation consisting of the absence of a graphic symbol. It allows automatic line break after if not followed by another SPACE.

NO-BREAK SPACE (00A0): NO-BREAK SPACE (NBSP) is a graphic character, the visual representation of which consists of the absence of a graphic symbol, for use when an automatic line break just before or just after it is to be prevented in the text as presented.

HYPHEN-MINUS (002D): HYPHEN-MINUS allows an automatic line break to be established just after it only if it is both immediately preceded by a letter and immediately followed by a letter. HYPHEN-MINUS should be imaged by a graphic symbol identical with that representing HYPHEN when immediately preceded or immediately followed by a letter. HYPHEN-MINUS should be imaged by a graphic symbol identical with that representing MINUS otherwise.

HYPHEN (2010): HYPHEN allows an automatic line break to be established just after it. HYPHEN is imaged by a graphic symbol.

NON-BREAKING HYPHEN (2011): NON-BREAKING HYPHEN is a graphic character, the visual representation of which is identical to that of HYPHEN. NON-BREAKING HYPHEN is for use as hyphen when an automatic line break just before or just after it is to be prevented in the text as presented.

SOFT HYPHEN (00AD): SOFT HYPHEN (SHY) allows an automatic line break to be established just after it (like ZERO WIDTH SPACE). SOFT HYPHEN is imaged by a graphic symbol identical with that representing HYPHEN when an automatic line break has been established just after it, or if it is directly followed by an explicit line break (including end-of-string). When an automatic line break has not been established just after it, nor is it followed by an explicit line break, the SOFT HYPHEN is not rendered and has zero width.

Note: In certain combinations, e.g., *webb<SHY>besökare*, the SOFT HYPHEN can in addition suppress the letter following the SOFT HYPHEN when the SOFT HYPHEN is not rendered (e.g. *webbesökare*). Such behaviour is similar to automatic ligature formation.

MONGOLIAN TODO SOFT HYPHEN (1806): MONGOLIAN TODO SOFT HYPHEN allows an automatic line break to be established just before it. MONGOLIAN TODO SOFT HYPHEN is imaged by a graphic symbol identical with that representing HYPHEN when an automatic line break has been established just before it, or if it is directly preceded by an explicit line break (including beginning-of-string). When an automatic line break has not been established just before it, nor is it preceded by an explicit line break, the MONGOLIAN TODO SOFT HYPHEN is not rendered and has zero width.